

RESEARCH

Open Access



Using radiomics for predicting the HPV status of oropharyngeal tumors

Kubra Sarac^{1*}  and Albert Guvenis¹

*Correspondence:
kubra.sarac@boun.edu.tr

¹ Institute of Biomedical
Engineering, Bogazici University,
Istanbul, Turkey

Abstract

Knowing human papillomavirus (HPV) status has important consequences for treatment selection in oropharyngeal cancer. The gold standard is to perform a biopsy. The objective of this paper is to develop a new computed tomography (CT) radiomics-based non-invasive solution to HPV status determination and investigate if and how it can be a viable and accurate complementary technique. Two hundred thirty-eight patients' CT scans were normalized and resampled. One thousand one hundred forty-two radiomics features were obtained from the segmented CT scans. The number of radiomic attributes was decreased by applying correlation coefficient analysis, backward elimination, and random forest feature importance analysis. Random over-sampling (ROSE) resampling algorithm was performed on the training set for data balancing, and as a result, 161 samples were obtained for each of the HPV classes of the training set. A random forest (RF) classification algorithm was used as a prediction model using five-fold cross-validation (CV). Model effectiveness was evaluated on the unused 20% of the imbalanced data. The applicability of the model was investigated based on previous research and error rates reported for biopsy procedures. The HPV status was determined with an accuracy of 91% (95% CI 83–99) and an area under the curve (AUC) of 0.77 (95% CI 65–89) on the test data. The error rates were comparable to those encountered in biopsy. As a conclusion, radiomics has the potential to predict HPV status with accuracy levels that are comparable to biopsy. Future work is needed to improve standardization, interpretability, robustness, and reproducibility before clinical translation.

Keywords: Radiomics, HPV status, Machine learning, Computed tomography, Oropharyngeal cancer

Introduction

The seventh most prevalent type of cancer worldwide and the ninth most common type of cancer in the USA, head and neck cancer (HNC), refer to a variety of upper aerodigestive tract tumors [1]. Around 644,000 new HNC cases are projected to be diagnosed annually worldwide, with two thirds of these occurrences taking place in the developing nations [2]. The American Joint Committee on Cancer defines HNC as a tumor

originating from both major and minor salivary glands as well as malignancies coming from mucosal areas of the oral cavity, larynx, paranasal sinuses, and pharynx [3]. Important risk factors for HNC include smoking drinking alcohol, being overexposed to sunlight, gamma, and ultraviolet radiation, having cancer in the family [4]. Additionally, human papillomavirus (HPV) has been linked to oropharyngeal cancer (OPC), which is a type of HNC. This type of cancer linked to HPV makes up approximately 25% of all HNCs [5]. The National Comprehensive Cancer Network (NCCN) recommends HPV testing for all oropharyngeal tumors in their guidelines [6]. In the USA, the percentage of head and neck cancers diagnosed as OPC that tested positive for HPV increased from 16.3% in the 1980s to more than 72.7% in the 2000s. This seems to be a result of increased awareness, the discovery of the link between HPV and cancers of the head and neck, and improved diagnostic HPV testing [7].

Importance of knowing the HPV status

Planning a course of treatment requires knowledge of the HPV status in OPC patients. HPV-positive OPC has a lower mortality rate than HPV-negative illness, with a 60% mortality rate with N3 or M1 sickness and an 80–90% 5-year survival rate even with lymph node involvement. Both overall mortality from all causes (10.4% vs. 33.3%) and mortality primarily from head and neck cancer (4.8% vs. 16.2%) are lower for HPV-positive patients. Despite having a poor 5-year survival rate of about 67%, OPSCC, which tests negative for HPV, has a poor prognosis [8, 9]. Therefore, knowing a patient's HPV status aids in identifying those who may have a better prognosis and may not need an aggressive course of therapy. According to studies, radiation and chemotherapy treatments have a stronger tendency to decrease and control HPV-positive tumors' growth. Contrarily, HPV-negative cancers might be more resistant to conventional therapies, requiring more intensive or unconventional therapeutic modalities [10]. Additionally, there is significant interest in de-escalating treatment intensity for patients with HPV-positive oropharyngeal cancer in order to reduce treatment-related toxicities while preserving outstanding results due to the favorable prognosis of this kind of disease [11]. Identifying patients who may be candidates for treatment de-escalation methods, such as lowering radiation dosage or chemotherapy intensity, is made easier with the use of HPV status information. This strategy aims to achieve the best possible compromise between reducing adverse effects from the treatment and controlling the tumor effectively [12]. Eligibility for particular clinical trials and targeted therapy is influenced by HPV status in oropharyngeal cancer. For individuals who are HPV-positive, several studies and cutting-edge treatments explicitly targeting HPV-related biological pathways may be beneficial. Clinicians can find suitable clinical trial choices and possibly investigate targeted treatments based on the underlying genetic features of the tumor by determining the HPV status [13]. The presence of HPV may also affect post-treatment surveillance plans. A more targeted and unique approach to post-treatment monitoring is made possible by modifying the surveillance procedures based on HPV status [14]. In the end, determining HPV status enhances the accuracy of therapy selection and helps to improve patient outcomes in oropharyngeal cancer.

Diagnostic and characterization methods

For oropharyngeal cancer, a variety of diagnostic and characterization techniques are currently available and in use [15]. Rapid diagnosis and treatment increase a patient's likelihood of recovering from an illness [16]. Oropharyngeal tumors are routinely diagnosed and characterized using a variety of traditional techniques, such as physical examination, imaging studies, and tissue samples [17]. A physical examination will be performed, during which a medical practitioner will carefully inspect the head, neck, and oropharynx. Any abnormal growths or other symptoms that might point to the presence of a tumor will be examined by medical professionals [18]. The oropharyngeal region can be seen, and tumors can be detected using a variety of imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and endoscopy [19]. For establishing the existence of a tumor and identifying its features, biopsy tissue sampling is essential. An oropharynx biopsy entails the removal of a tiny sample of tissue, which is then sent to a lab for examination. HPV molecular testing is possible on this tissue [20]. Conventional laboratory techniques like p16 immunohistochemistry (IHC) and polymerase chain reaction (PCR) can be utilized to determine the HPV status of a patient. The 8th edition of the American Joint Commission on Cancer (AJCC) recommended P16 IHC as a diagnostic test for oropharyngeal cancer staging [21]. However, this may increase the effort and cause a delay in clinical applications. It is crucial to remember that the diagnostic procedure can change based on the particular instance, the preferences of the healthcare professional, and the accessibility of resources. To guarantee a precise diagnosis and thorough understanding of oropharyngeal tumors, a multidisciplinary strategy comprising numerous specialists, such as otolaryngologists, radiologists, and pathologists, is frequently used.

A new alternative: radiomics

In addition to all of the classic diagnostic techniques mentioned above, new technique known as radiomics has recently attracted interest. The field of radiomics has advanced swiftly toward practical application in the hopes of improving cancer treatment and accurate detection. Radiomics is a quantitative approach to medical imaging that makes use of sophisticated mathematical analysis to enhance the data already available to doctors [22]. Radiomics quantifies textural information by mathematically extracting the spatial distribution of signal intensities and pixel interrelationships using analysis methods from the field of artificial intelligence [23]. Due to their potential prognostic value for treatment outcomes radiomic features have recently gained a lot of attention and may be useful in personalized medicine.

The application of radiomics in oropharyngeal malignancy diagnosis and precision medicine has shown promise since it enhances characterization and diagnosis. This improved characterization can help in making the distinction between benign and malignant tumors, determining how aggressive the tumor is, and guiding treatment choices [22]. Radiation oncologists can optimize radiation dose distribution and intensity modulation to target the tumor more precisely while preserving healthy tissues by incorporating radiomic characteristics into treatment planning algorithms. Radiation therapy when guided by radiomics attempts to provide a more individualized type of care that maximizes tumor control while minimizing adverse effects [24].

Radiomics can also help in calculating the likelihood of metastasis, disease recurrence, or overall survival, which can inform treatment choices and follow-up plans [25].

In some circumstances, radiomics may be used in addition to or in very exceptional cases as a substitute for biopsy [26]. Radiomics can also be used to track a cancer patient's reaction to treatment. The success of treatment and tumor development or regression may be determined by tracking changes in radiomic characteristics over time [22]. In summary, the radiomics technique has the ability to capture complex tumor properties, in addition to what can be learned from a biopsy sample alone [26].

By identifying the most questionable or important locations for the sample, radiology can help direct the biopsy procedure. Radiomics can increase the precision and diagnostic yield of biopsies by examining radiomic characteristics, imaging-based biomarkers, or tumor heterogeneity patterns, ensuring that the most representative tissue samples are obtained [27].

Biopsy and radiomics are two different methods that have advantages and drawbacks. Radiomics' non-invasive nature is undoubtedly one of its advantages. Both geographically and temporally diverse tumors exist in solid tumors. As a result, the use of invasive biopsy-based molecular assays is constrained.

Radiomics, which may non-invasively detect intra-tumoral heterogeneity, now has a wide range of applications [28].

In summary, radiomics provides a non-invasive whole-lesion assessment with the possibility of temporal monitoring and multi-dimensional analysis. However, it is deficient in tissue confirmation and thorough histological data.

Contribution

In the literature, a number of algorithms have been proposed for radiomics-based HPV status determination [29–41]. This work is different in that the developed method is based on a not-yet-tried machine-learning algorithm combination. Additionally, a thorough comparison of our approach and results with the biopsy technique was conducted for the first time.

The advantages and disadvantages of each approach and typical error rates encountered are given. Finally, the future work needed to translate the non-invasive radiomics approach to routine clinical practice is outlined.

Material and methods

Data set

Four hundred ninety-five patients in the collection of data from The Cancer Imaging Archive (TCIA) [42] were included in the study. The study only examined 238 individuals' contrast-enhanced CT scans, 204 of whom were HPV positive and had been given an OPC diagnosis. The gross primary tumor volume (GTVp) (see Fig. 1), which is divided by experts, is taken into account in radiomic research [35].

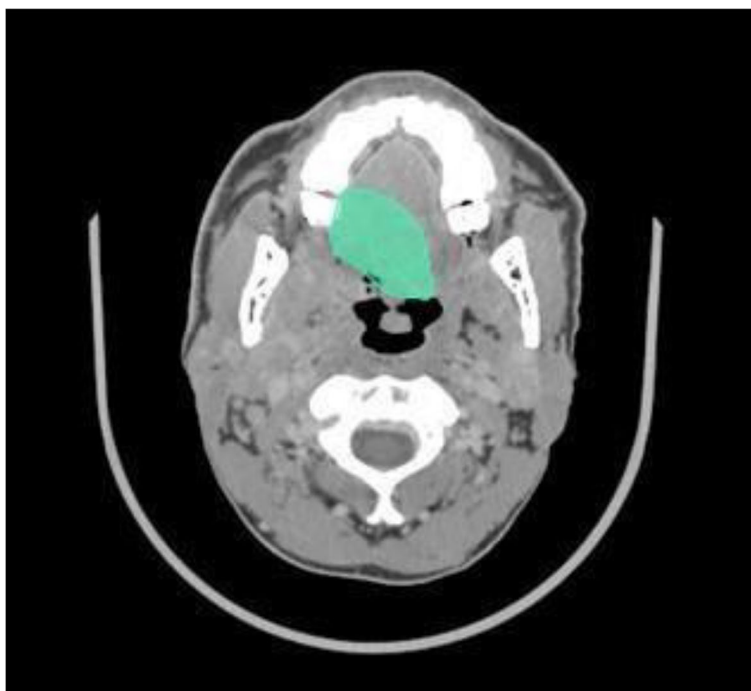


Fig. 1 One slice of CT image of a patient with OPC. Green-colored segmentation represents the tumor area

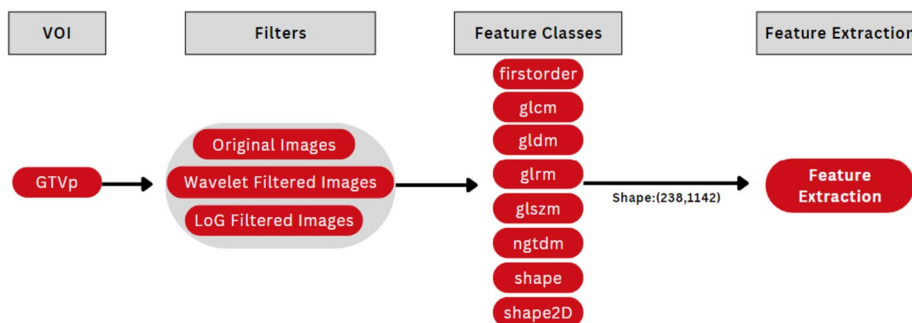


Fig. 2 Feature extraction process

Image pre-processing

Prior to the radiomic feature extraction procedure, all patients’ CT images were resampled and normalized. Resampling CT images produced 1 mm × 1 mm × 1 mm voxels [43]. The three-dimensional (3D) slicer program (version 5.0.3) was used to do resampling and interpolation methods using the Python-based pyradiomics [44].

Feature extraction

After performing image preprocessing, the feature extraction process was carried out using 3D Slicer software (version 5.0.3) [45]. A standard bin width of 10 was set in order to implement gray-level discretization and reduce variability [46]. These

included characteristics of the original images, wavelet-transformed images, and Laplacian of Gaussian (Log)-filtered images (see Fig. 2). After the images have been converted into features, the newly generated data were utilized to train and evaluate machine learning (ML) models. These characteristics can be used to perform quantitative image comparisons [47]. van Griethuysen et al. [48] have detailed explanations of the radiomics technique.

Data pre-processing and resampling

The 1142 features were subject to Z-score normalization. Twenty percent of the data was designated for testing, while the remaining 80% was designated for training. There was an uneven distribution of HPV classes between the training set and the test set. The test set contained 48 instances (43 HPV positive, 5 HPV negative), while the training set contained 190 cases (161 HPV positive, 29 HPV negative).

Since the number of cases with HPV status was imbalanced, the random over-sampling (ROSE) [49] resampling method was utilized. Only the training set was subjected to a resampling technique, and 161 samples for each positive and negative HPV class were obtained (see Fig. 3).

Feature selection

Radiomics approaches typically produce high dimensional data, which increases the risk of over-fitting, worsens model confusion, and degrades prediction accuracy. In this research, correlation coefficient analysis (CCA), random forest (RF) feature importance analysis, and backward elimination methods were used in order to choose functional features (see Fig. 4). CCA was initially employed as a filter-based technique to subtract unneeded characteristics that were extremely closely related (absolute correlation coefficient > 0.9). The Gini impurity metric, which offers a better way to gauge feature importance, was used to run the random forest model [50].

The fifty most crucial traits were chosen using the sequential backward selection approach, with the k-nearest neighbor serving as a forecaster. The feature selection

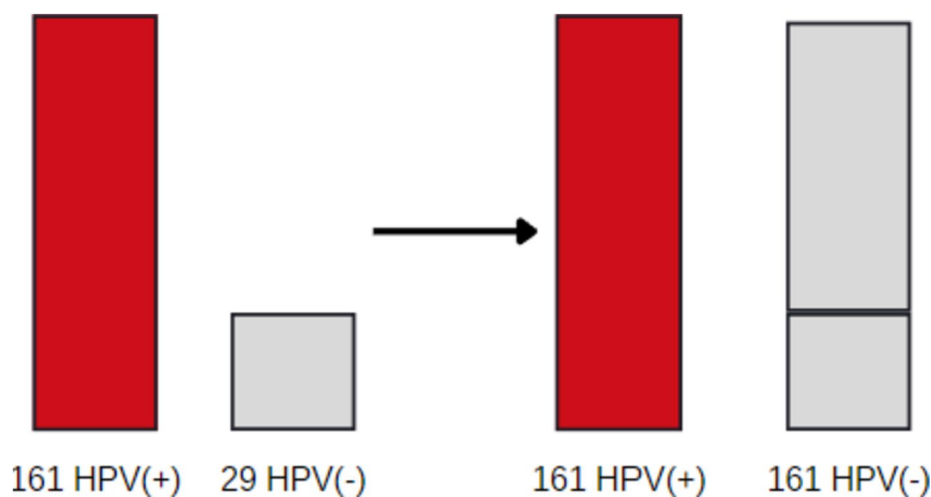


Fig. 3 Random over-sampling application on training data

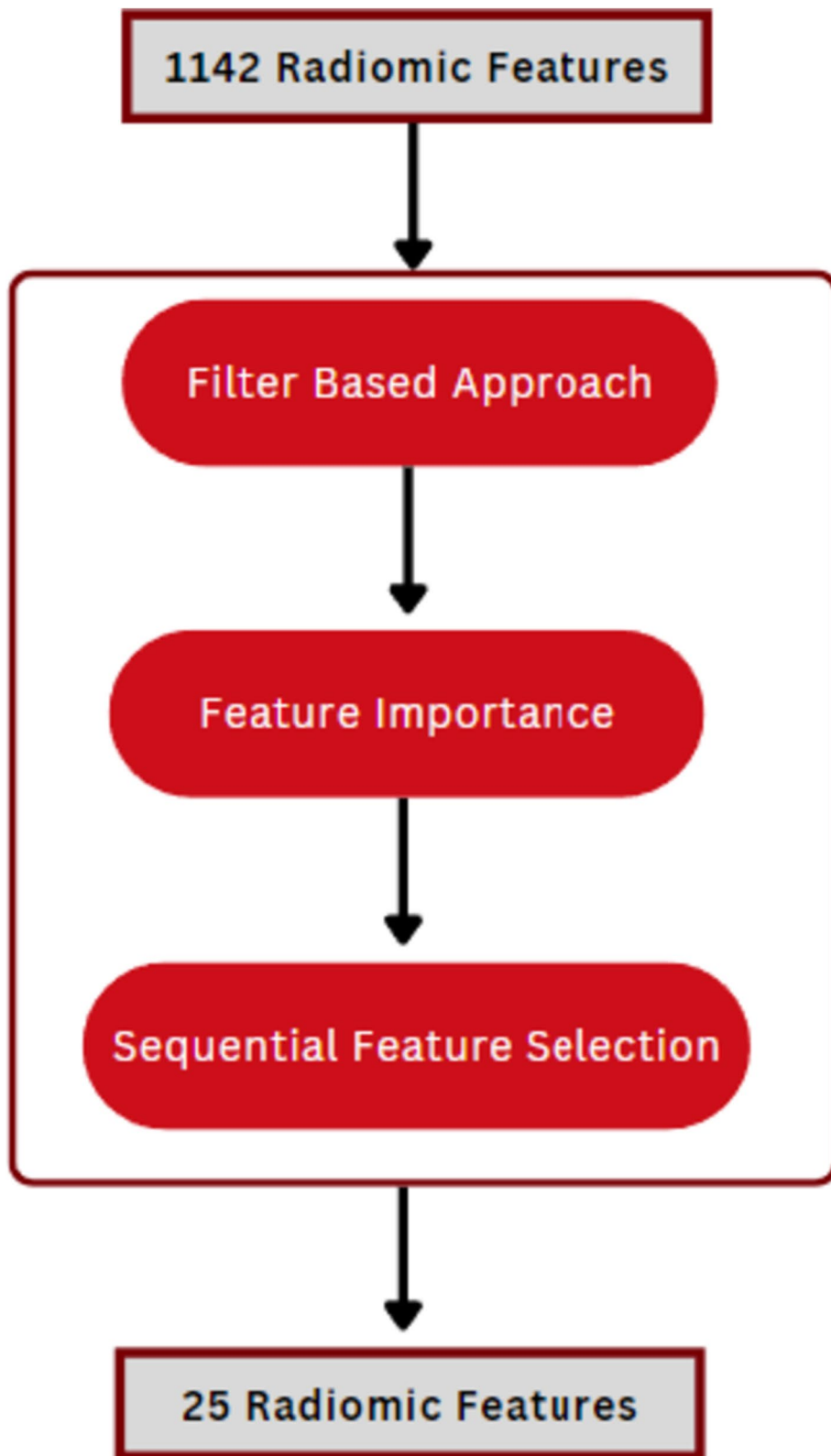


Fig. 4 Flowchart for the feature selection process

techniques were carried out in Python (version 3.9) using the MLxtend and Scikit-learn libraries [51].

Results

Model training and evaluation

A random forest (RF) classification ML algorithm was used as a prediction ML model utilizing five-fold cross-validation (CV). Five hundred different combinations of hyper-parameters were tried on the RF model utilizing the randomized search CV method on training feature sets to determine the best ones that maximize model efficiency. This operation (five-fold nested cross-validation) was run five times for various training and testing sets. On the 20% of the initial unbalanced data that was not used, the model’s performance was assessed separately. HPV status was predicted by an RF algorithm with an accuracy of 91% (95% CI 83–99) and an area under the curve (AUC) of 0.77 (95% CI 65–89) on the test data. The confusion matrix and ROC curve (receiver operating characteristic curve) for the random forest model with ROSE resampling algorithm performance result on the test data are demonstrated in Figs. 5 and 6, respectively.

Discussion

The objective of this paper was to develop a new radiomics-based solution to the problem of HPV determination for OPC patients and make a thorough analysis of its applicability in routine clinical practice.

The RF algorithm in combination with resampling allowed us to identify the HPV situation with an accuracy of 91% (95% CI 83–99) AUC of 0.77 (95% CI 65–89) on the independent test data.

The RF algorithm has several known advantages such as good predictions that can be understood easily and a higher level of accuracy with respect to decision trees. It can

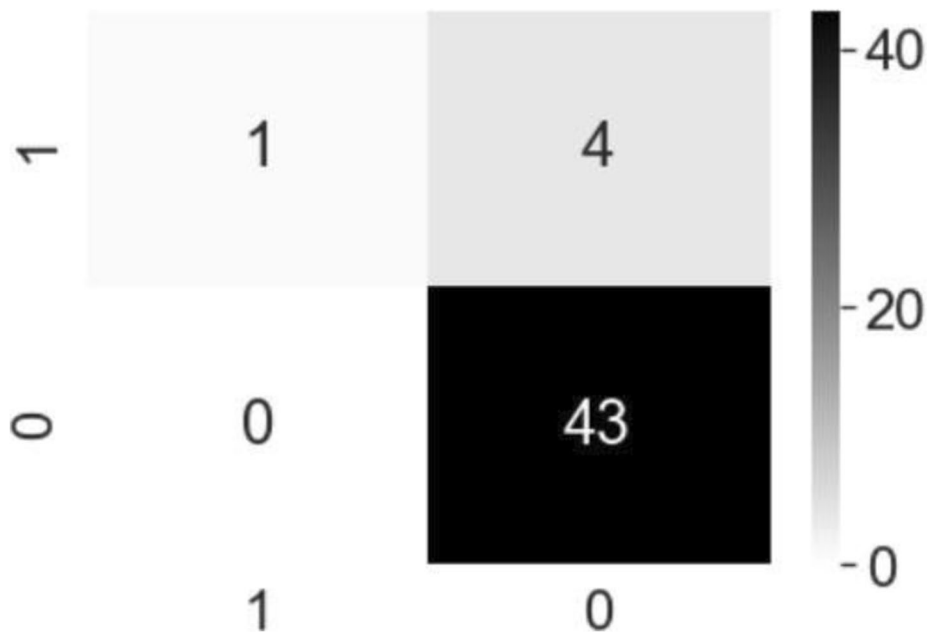


Fig. 5 Confusion matrix of random forest model with the ROSE re-sampling algorithm

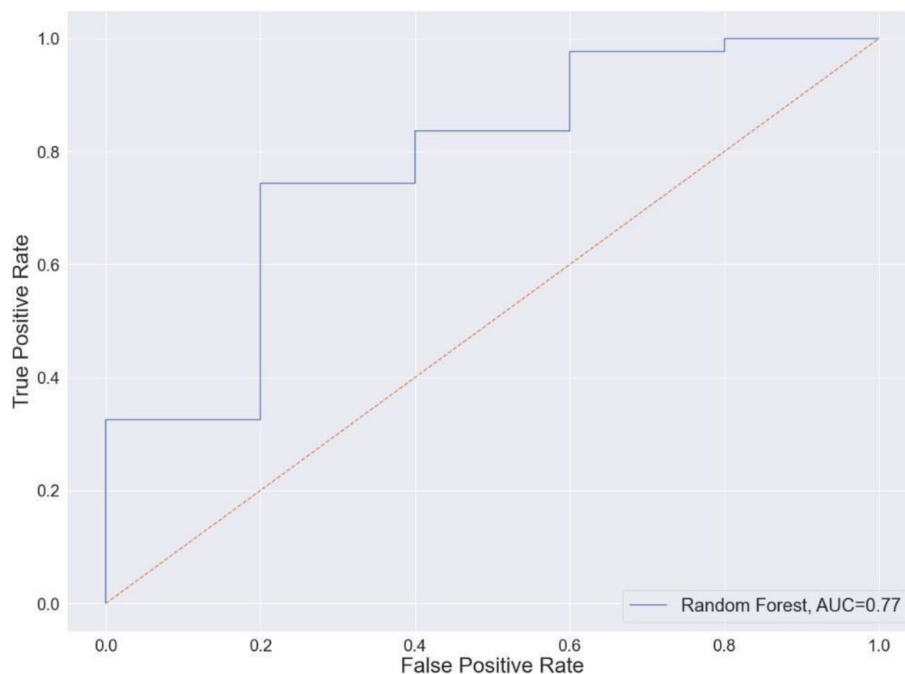


Fig. 6 ROC curve of random forest model with the ROSE re-sampling algorithm

also handle large datasets that may be available in the future. Other algorithms and different datasets have been investigated in several previous studies [29–41]. In the present study, the results were comparable to previous findings although the data size was small and highly imbalanced. Another limitation is that no testing has been done to determine the sensitivity of radiomic characteristics with respect to segmentation alterations. The results have been obtained using an independent test dataset but not have been verified with data obtained from other institutions.

Comparison with biopsy and future work for widespread application

Radiomics and biopsy can be complementary to each other. Some of the shortcomings of the biopsy are missing the most aggressive or representative areas of the tumor [52] and failing to detect cancer cells in the sample because the tumor can be small or located in a challenging anatomical site [53]. Furthermore, there can be inter-observer variability, where different pathologists may interpret the same biopsy sample differently, leading to variations in treatment decisions [54]. Biopsy samples may not fully capture the complexity of the tumor, including variations in genetic mutations, protein expression, or cellular characteristics. Besides, biopsies, especially those performed using invasive techniques such as surgical excision or fine needle aspiration, carry some risks and potential complications. These can include bleeding, infection, damage to nearby structures, and patient discomfort [55]. In [54], the sample error accounted for 60.0% of inconsistent findings, and pathologist inconsistency accounted for 23.3%. The error rates can change depending on the specific biopsy procedure and the condition being evaluated.

It is important to acknowledge these limitations when interpreting biopsy results in oropharyngeal cancer. Clinicians may want to consider a complementary approach such

as radiomics in order to obtain additional evidence, in particular when biopsy conditions are not optimal.

Radiomics has also its own disadvantages. There is a lack of standardized protocols and guidelines for feature extraction, leading to variability in the methods used across different studies and institutions. This lack of standardization can impact the reproducibility and comparability of radiomic results, making it challenging to establish consistent and reliable radiomic models. Radiomics heavily relies on the quality and consistency of the medical images used for analysis. However, imaging techniques, acquisition parameters, and equipment can vary between institutions, scanners, and even individual radiologists. These variations in image acquisition can introduce variability and bias in the radiomic features, affecting the accuracy and generalizability of the results. Moreover, radiomics relies on the availability of large and diverse datasets for training and validation purposes. However, obtaining high-quality imaging data with corresponding clinical annotations can be challenging due to issues such as data privacy, limited sample sizes, and variations in data collection across institutions. Limited data availability and potential biases in the data can impact the development and validation of robust radiomic models [56]. Furthermore, while radiomics studies have shown promising results in research settings, there is a need for robust validation and clinical translation. The performance of radiomic models in real-world clinical settings may differ from the initial research findings. Further validation studies, preferably in multi-institutional settings, are needed to establish the clinical usefulness and effectiveness of radiomics in various disease contexts. The radiomics results should also provide a probability for the likelihood of a correct result for a particular patient. Also, radiomic models often provide quantitative and statistical measures, but the interpretation of these measures and their integration into clinical decision-making can be challenging [27]. The clinical relevance and meaningfulness of radiomic features need to be further explored and validated to ensure their utility in guiding treatment decisions and patient management. Addressing these reported issues requires ongoing research and collaboration among radiomics researchers, imaging experts, and clinicians. The Image Biomarker Standardization Initiative (IBSI), founded by study participants [57], was created to overcome these difficulties by creating the goals, which are nomenclature and descriptions for frequently utilized radiomic characteristics; a common image processing using radiomics plan for the computation of imaging-based characteristics; and data collection and related reference values for the calibration and testing of image processing software implementations.

Future ML work

In spite that our models are able to forecast HPV status with a good level of AUC and relative accuracy, more research needs to be done utilizing larger clinical datasets to verify the effectiveness of the created ML model. All the above-mentioned concerns about radiomics should also be addressed (interpretability, standardization, reproducibility).

Conclusion

In conclusion, this work demonstrates that it is clinically important and possible to develop a new CT radiomic-based non-invasive complementary solution for the determination of HPV status with accuracy rates that can challenge those obtained from biopsy. However, further research is needed to improve accuracy, safety, standardization, interpretability, and reproducibility for widespread clinical acceptance.

Abbreviations

3D	Three-dimensional
AJCC	American Joint Commission on Cancer
AUC	Area under the curve
CCA	Correlation coefficient analysis
CT	Computed tomography
CV	Cross-validation
EHR	Electronic health records
FNA	Fine needle aspiration
GLM	General linear model
GTVp	Gross primary tumor volume
HNC	Head and neck cancer
HNSCC	Head and neck squamous cell carcinoma
HPV	Human papillomavirus
IHC	Immunohistochemistry
LR	Logistic regression
Log	Laplacian of Gaussian
ML	Machine learning
MRI	Magnetic resonance imaging
NCCN	National Comprehensive Cancer Network
OPC	Oropharyngeal cancer
PACS	Picture Archiving and Communication Systems
PET	Positron emission tomography
PCR	Polymerase chain reaction
RF	Random forest
ROC	Receiver operating characteristic curve
ROI	Region of interest
ROSE	Random over-sampling
TCIA	The Cancer Imaging Archive
VOI	Volume of interest
Xgboost	Extreme gradient boosting

Acknowledgements

We would like to thank Oya Altinok for her valuable ideas during the study.

Authors' contributions

Study design, AG. Project management, AG. Implementation, KS. Manuscript preparation, KS. Review, AG.

Funding

This study was supported by Bogaziçi University Research Fund Grant Number 19703P.

Availability of data and materials

The data is available upon request to Researcher Kubra Sarac by email kubra.sarac@boun.edu.tr.

Declarations

Ethics approval and consent to participate

This study used a public dataset; no ethics approval is required.

Competing interests

The authors declare that they have no competing interests.

Received: 16 August 2023 Accepted: 22 December 2023

Published online: 04 January 2024

References

1. Rettig EM, D'Souza G (2015) Epidemiology of head and neck cancer. *Surg Oncol Clin N Am* 24:379–396. <https://doi.org/10.1016/j.soc.2015.03.001>

2. Marur S, Forastiere AA (2008) Head and neck cancer: changing epidemiology, diagnosis, and treatment. *Mayo Clin Proc* 83:489–501. <https://doi.org/10.4065/83.4.489>
3. Cohen N, Fedewa S, Chen AY (2018) Epidemiology and demographics of the head and neck cancer population. *Oral Maxillofac Surg Clin North Am* 30:381–395. <https://doi.org/10.1016/j.coms.2018.06.001>
4. Mahmood H, Shaban M, Rajpoot N, Khurram SA (2021) Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. *Br J Cancer* 124:1934–1940. <https://doi.org/10.1038/s41416-021-01386-x>
5. Tanaka TI, Alawi F (2018) Human papillomavirus and oropharyngeal cancer. *Dent Clin North Am* 62:111–120. <https://doi.org/10.1016/j.cden.2017.08.008>
6. Evaluating the incidence of HPV-positive/-negative, according to NCCN guidelines. <https://www.targetedonc.com/view/evaluating-the-incidence-of-hpv-positive--negative-according-to-nccn-guidelines>. Accessed 14 Dec 2023
7. Chow LQM (2020) Head and neck cancer. *N Engl J Med* 382(1):60–72. <https://doi.org/10.1056/nejma1715715>
8. Avery EW, Joshi K, Mehra S, Mahajan A (2023) Role of PET/CT in oropharyngeal cancers. *Cancers* 15(9):2651. <https://doi.org/10.3390/cancers15092651>
9. Rischin D, Young RJ, Fisher R, Fox SB, Le Q-T, Peters LJ, Solomon B, Choi J, O'Sullivan B, Kenny LM, McArthur GA (2010) Prognostic significance of p16ink4a and human papillomavirus in patients with oropharyngeal cancer treated on Trog 02.02 Phase III trial. *J Clin Oncol* 28(27):4142–4148. <https://doi.org/10.1200/jco.2010.29.2904>
10. Perri F, Longo F, Caponigro F, Sandomenico F, Guida A, Della Vittoria Scarpati G, Ottaiano A, Muto P, Ionna F (2020) Management of HPV-related squamous cell carcinoma of the head and neck: pitfalls and caveat. *Cancers* 12(4):975. <https://doi.org/10.3390/cancers12040975>
11. Zakeri K, Dunn L, Lee N (2021) HPV-associated oropharyngeal cancer de-escalation strategies and trials: past failures and future promise. *J Surg Oncol* 124(6):962–966. <https://doi.org/10.1002/jso.26696>
12. Kimple RJ, Harari PM (2014) Is radiation dose reduction the right answer for HPV-positive head and neck cancer? *Oral Oncol* 50(6):560–564. <https://doi.org/10.1016/j.oraloncology.2013.09.015>
13. Bonilla-Velez J, Mroz EA, Hammon RJ, Rocco JW (2013) Impact of human papillomavirus on oropharyngeal cancer biology and response to therapy. *Otolaryngol Clin North Am* 46(4):521–543. <https://doi.org/10.1016/j.otc.2013.04.009>
14. Dermody, S. M., Haring, C. T., Bhambhani, C., Tewari, M., Brenner, J. C., & Swiecicki, P. L. (2021). Surveillance and monitoring techniques for HPV-related head and neck squamous cell carcinoma: circulating tumor DNA. *Current Treatment Options in Oncology*, 22(3). <https://doi.org/10.1007/s11864-021-00821-8>
15. Alam S, Chaurasia A, Singh N (2021) Oral cancer diagnostics: an overview. *Nat J Maxillofacial Surg* 12(3):324. https://doi.org/10.4103/njms.njms_130_20
16. Sciubba JJ (2001) Oral cancer. *Am J Clin Dermatol* 2(4):239–251. <https://doi.org/10.2165/00128071-200102040-00005>
17. Macey, R., Walsh, T., Brocklehurst, P., Kerr, A. R., Liu, J. L., Lingen, M. W., Ogden, G. R., Warnakulasuriya, S., & Scully, C. (2015). Diagnostic tests for oral cancer and potentially malignant disorders in patients presenting with clinically evident lesions. *Cochrane Database Syst Rev*. <https://doi.org/10.1002/14651858.cd010276.pub2>
18. Cohan DM, Popat S, Kaplan SE, Rigual N, Loree T, Hicks WL Jr (2009) Oropharyngeal cancer: current understanding and management. *Curr Opin Otolaryngol Head Neck Surg* 17(2):88–94. <https://doi.org/10.1097/moo.0b013e32832984c0>
19. Tshering Vogel, D. W., Zbaeren, P., & Thoeny, H. C. (2010). Cancer of the oral cavity and oropharynx. *Cancer Imaging*, 10(1). <https://doi.org/10.1102/1470-7330.2010.0008>
20. Yang G, Wei L, Thong BK, Fu Y, Cheong IH, Kozlakidis Z, Li X, Wang H, Li X (2022) A systematic review of oral biopsies, sample types, and detection techniques applied in relation to oral cancer detection. *Biotech* 11:5. <https://doi.org/10.3390/biotech11010005>
21. Hoffmann M, Tribius S (2019) HPV and oropharyngeal cancer in the eighth edition of the TNM classification: pitfalls in practice. *Transl Oncol* 12:1108–1112. <https://doi.org/10.1016/j.tranon.2019.05.009>
22. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J (2019) The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics* 9:1303–1322. <https://doi.org/10.7150/thno.30309>
23. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhri H, Baessler B (2020) Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights into Imaging*. <https://doi.org/10.1186/s13244-020-00887-2>
24. Abdollahi H, Chin E, Clark H, Hyde DE, Thomas S, Wu J, Uribe CF, Rahmim A (2022) Radiomics-guided radiation therapy: opportunities and challenges. *Phys Med Biol*. <https://doi.org/10.1088/1361-6560/ac6fab>
25. Rich B, Huang J, Yang Y, Jin W, Johnson P, Wang L, Yang F (2021) Radiomics predicts for distant metastasis in locally advanced human papillomavirus-positive oropharyngeal squamous cell carcinoma. *Cancers* 13:5689. <https://doi.org/10.3390/cancers13225689>
26. Scheckenbach K (2018) Radiomics: big data Statt Biopsie in der Zukunft? *Laryngo-Rhino-Otologie*. <https://doi.org/10.1055/s-0043-121964>
27. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577. <https://doi.org/10.1148/radiol.2015151169>
28. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL (2012) Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48:441–446. <https://doi.org/10.1016/j.ejca.2011.11.036>
29. Tortora M, Gemini L, Scaravilli A, Ugga L, Ponsiglione A, Stanzione A, D'Arco F, D'Anna G, Cuocolo R (2023) Radiomics applications in head and neck tumor imaging: a narrative review. *Cancers* 15:1174. <https://doi.org/10.3390/cancers15041174>
30. Spadarella G, Ugga L, Calareso G, Villa R, D'Aniello S, Cuocolo R (2022) The impact of radiomics for human papillomavirus status prediction in oropharyngeal cancer: systematic review and Radiomics Quality Score Assessment. *Neuroradiology* 64:1639–1647. <https://doi.org/10.1007/s00234-022-02959-0>

31. Bos P, Brekel MW, Gouw ZA, Al-Mamgani A, Waktola S, Aerts HJ, Beets-Tan RG, Castelijns JA, Jasperse B (2020) Clinical variables and magnetic resonance imaging-based radiomics predict human papillomavirus status of oropharyngeal cancer. *Head Neck* 43:485–495. <https://doi.org/10.1002/hed.26505>
32. Zhinan L, Wei Z, Yudi Y, Yabing D, Yuanzhe X, Xiulan L (2022) Prediction of HPV status in oropharyngeal squamous cell carcinoma based on radiomics and machine learning algorithms: a multi-cohort study. <https://doi.org/10.21203/rs.3.rs-1841205/v1>
33. Suh CH, Lee KH, Choi YJ, Chung SR, Baek JH, Lee JH, Yun J, Ham S, Kim N (2020) Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametricmr images for determination of HPV infection status. *Sci Rep*. <https://doi.org/10.1038/s41598-020-74479-x>
34. Boot PA, Mes SW, de Bloeme CM, Martens RM, Leemans CR, Boellaard R, van de Wiel MA, de Graaf P (2023) Magnetic resonance imaging based radiomics prediction of human papillomavirus infection status and overall survival in oropharyngeal squamous cell carcinoma. *Oral Oncol* 137:106307. <https://doi.org/10.1016/j.oraloncology.2023.106307>
35. Song B, Yang K, Garneau J, Lu C, Li L, Lee J, Stock S, Braman NM, Koyuncu CF, Toro P, Fu P, Koyfman SA, Lewis JS, Madabhushi A (2021) Radiomic features associated with HPV status on pretreatment computed tomography in oropharyngeal squamous cell carcinoma inform clinical prognosis. *Front Oncol*. <https://doi.org/10.3389/fonc.2021.744250>
36. Altinok O, Guvenis A (2022) Interpretable radiomics method for predicting human papillomavirus status in oropharyngeal cancer using Bayesian networks. <https://doi.org/10.1101/2022.06.29.22276890>
37. Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, Guckenberger M, Tanadini-Lang S (2017) 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* 99:921–928. <https://doi.org/10.1016/j.ijrobp.2017.06.002>
38. Bagher-Ebadian H, Lu M, Siddiqui F, Ghanem AI, Wen N, Wu Q, Liu C, Movsas B, Chetty IJ (2020) Application of radiomics for the prediction of HPV status for patients with head and neck cancers. *Med Phys* 47:563–575. <https://doi.org/10.1002/mp.13977>
39. Yu K, Zhang Y, Yu Y, Huang C, Liu R, Li T, Yang L, Morris JS, Baladandayuthapani V, Zhu H (2017) Radiomic analysis in prediction of human papilloma virus status. *Clin Transl Radiat Oncol* 7:49–54. <https://doi.org/10.1016/j.ctro.2017.10.001>
40. Reiazi R, Arrowsmith C, Welch M, Abbas-Aghababazadeh F, Eeles C, Tadic T, Hope AJ, Bratman SV, Haibe-Kains B (2021) Prediction of human papillomavirus (HPV) Association of Oropharyngeal Cancer (OPC) using radiomics: the impact of the variation of CT scanner. *Cancers* 13:2269. <https://doi.org/10.3390/cancers13092269>
41. Sarac K, Guvenis A (2023) Determining HPV status in patients with oropharyngeal cancer from 3D CT images using radiomics: effect of sampling methods. *Bioinform Biomed Eng* 27–41. https://doi.org/10.1007/978-3-031-34960-7_3
42. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F (2013) The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging* 26:1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
43. Karagöz A, Guvenis A (2022) Robust whole-tumour 3D volumetric CT-based radiomics approach for predicting the WHO/ISUP grade of a CCRCC tumour. *Comput Methods Biomech Biomed Eng* 11:665–677. <https://doi.org/10.1080/21681163.2022.2103449>
44. Wels MG, Lades F, Muehlberg A, Suehling M (2019) General purpose radiomics for multi-modal clinical research. *Medical Imaging 2019: Computer-Aided Diagnosis*. <https://doi.org/10.1117/12.2511856>
45. Chianca V, Cuocolo R, Gitto S, Albano D, Merli I, Badalyan J, Cortese MC, Messina C, Luzzati A, Parafioriti A, Galbusera F, Brunetti A, Sconfienza LM (2021) Radiomic machine learning classifiers in spine bone tumors: a multi-software, multi-scanner study. *Eur J Radiol* 137:109586. <https://doi.org/10.1016/j.ejrad.2021.109586>
46. Larue RT, van Timmeren JE, de Jong EE, Feliciani G, Leijenaar RT, Schreurs WM, Sosef MN, Raat FH, van der Zande FH, Das M, van Elmpt W, Lambin P (2017) Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol* 56:1544–1553. <https://doi.org/10.1080/0284186x.2017.1351624>
47. Tamal M (2019) Grey level co-occurrence matrix (GLCM) as a Radiomics feature for artificial intelligence (AI) assisted positron emission tomography (PET) images analysis. *IOP Conference Series Mater Sci Eng* 646:012047. <https://doi.org/10.1088/1757-899x/646/1/012047>
48. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. <https://doi.org/10.1158/0008-5472.can-17-0339>
49. Mohammed R, Rawashdeh J, Abdullah M (2020) Machine learning with oversampling and undersampling techniques: overview study and experimental results. 2020 11th International Conference on Information and Communication Systems (ICICS). <https://doi.org/10.1109/icics49469.2020.239556>
50. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA (2009) A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of Spectral Data. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-10-213>
51. Stancin I, Jovic A (2019) An overview and comparison of free python libraries for data mining and big data analysis. 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). <https://doi.org/10.23919/mipro.2019.8757088>
52. Lange MB, Petersen LJ, Nielsen MB, Zacho HD (2021) Validity of negative bone biopsy in suspicious bone lesions. *Acta Radiologica Open* 10:205846012110306. <https://doi.org/10.1177/20584601211030662>
53. Göret CC, Göret NE, Özdemir ZT, Özkan EA, Doğan M, Yanık S, Gümrükçü G, Aker FV (2015) Diagnostic value of fine needle aspiration biopsy in non-thyroidal head and neck lesions: a retrospective study of 866 aspiration materials. *Int J Clin Exp Pathol* 8(8):8709–8716
54. Chen S, Forman M, Sadow PM, August M (2016) The diagnostic accuracy of incisional biopsy in the oral cavity. *J Oral Maxillofac Surg* 74:959–964. <https://doi.org/10.1016/j.joms.2015.11.006>
55. S; SD Fine needle aspiration. In: National Center for Biotechnology Information. <https://pubmed.ncbi.nlm.nih.gov/32491418/>. Accessed 14 Aug 2023

56. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpst W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
57. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook GJ, Davatzikos C, Depeursinge A, Desseroit M-C, Dinapoli N, Dinh CV, Echegaray S, El Naqa I, Fedorov AY, Gatta R, Gillies RJ, Goh V, Götz M, Guckenberger M, Ha SM, Hatt M, Isensee F, Lambin P, Leger S, Leijenaar RTH, Lenkowicz J, Lippert F, Losnegård A, Maier-Hein KH, Morin O, Müller H, Napel S, Nioche C, Orhac F, Pati S, Pfaehler EAG, Rahmim A, Rao AUK, Scherer J, Siddique MM, Sijtsema NM, Socarras Fernandez J, Spezi E, Steenbakkens RJHM, Tanadini-Lang S, Thorwarth D, Troost EGC, Upadhaya T, Valentini V, van Dijk LV, van Griethuysen J, van Velden FHP, Whybra P, Richter C, Löck S (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295:328–338. <https://doi.org/10.1148/radiol.2020191145>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
