

Article

A Radiomics-Based Classifier for the Progression of Oropharyngeal Cancer Treated with Definitive Radiotherapy

Darwin A. Garcia ^{1,2,*} , Elizabeth B. Jeans ¹, Lindsay K. Morris ¹, Satomi Shiraishi ¹, Brady S. Laughlin ³ , Yi Rong ³, Jean-Claude M. Rwigema ³, Robert L. Foote ¹ , Michael G. Herman ¹ and Jing Qian ^{1,*}

¹ Department of Radiation Oncology, Mayo Clinic, Rochester, MN 55905, USA

² Mayo Clinic Graduate School of Biomedical Sciences, Mayo Clinic, Rochester, MN 55905, USA

³ Department of Radiation Oncology, Mayo Clinic, Phoenix, AZ 85054, USA

* Correspondence: garcia.darwin@mayo.edu (D.A.G.); qian.jing@mayo.edu (J.Q.)

Simple Summary: Oropharyngeal cancer is the most common type of head-and-neck squamous cell carcinoma. Although patients with HPV-associated cancers have a better prognosis than patients with HPV-negative cancers, there is a lack of robust biomarkers that describe the relative risk for disease progression. To investigate this problem, we extracted quantitative descriptors from medical images (known as radiomics features) that could not otherwise be assessed by the naked eye. We built a machine learning model based on clinical and radiomics features to predict whether a patient will exhibit disease progression at 2 years post-treatment. These findings are important for identifying patients treated with definitive radiotherapy with low and high risk of disease progression and formulating patient-specific treatment strategies.

Abstract: In this study, we investigated whether radiomics features from pre-treatment positron emission tomography (PET) images could be used to predict disease progression in patients with HPV-positive oropharyngeal cancer treated with definitive proton or x-ray radiotherapy. Machine learning models were built using a dataset from Mayo Clinic, Rochester, Minnesota ($n = 72$) and tested on a dataset from Mayo Clinic, Phoenix, Arizona ($n = 22$). A total of 71 clinical and radiomics features were considered. The Mann–Whitney U test was used to identify the top 2 clinical and top 20 radiomics features that were significantly different between progression and progression-free patients. Two dimensionality reduction methods were used to define two feature sets (manually filtered or machine-driven). A forward feature selection scheme was conducted on each feature set to build models of increased complexity (number of input features from 1 to 6) and evaluate model robustness and overfitting. The machine-driven features had superior performance and were less prone to overfitting compared to the manually filtered features. The four-variable Gaussian Naïve Bayes model using the ‘Radiation Type’ clinical feature and three machine-driven features achieved a training accuracy of 79% and testing accuracy of 77%. These results demonstrate that radiomics features can provide risk stratification beyond HPV-status to formulate individualized treatment and follow-up strategies.

Keywords: PET; radiomics; oropharyngeal cancer; definitive radiotherapy; progression



Citation: Garcia, D.A.; Jeans, E.B.; Morris, L.K.; Shiraishi, S.; Laughlin, B.S.; Rong, Y.; Rwigema, J.-C.M.; Foote, R.L.; Herman, M.G.; Qian, J. A Radiomics-Based Classifier for the Progression of Oropharyngeal Cancer Treated with Definitive Radiotherapy. *Cancers* **2023**, *15*, 3715. <https://doi.org/10.3390/cancers15143715>

Academic Editor: David Wong

Received: 19 June 2023

Revised: 14 July 2023

Accepted: 20 July 2023

Published: 22 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Oropharyngeal cancer (OPC) is the most prevalent type of head-and-neck squamous cell carcinoma. Despite a decline in the incidence of head-and-neck-squamous cell cancers in recent decades [1], human papillomavirus (HPV) has driven an increase in the incidence of OPC across various demographic groups [2,3].

HPV-positive OPC is staged as a disease entity that is distinct from HPV-negative OPC, with favorable prognosis associated with HPV-positive cases [4,5]. Patients with HPV-positive OPC are younger at diagnosis, have fewer medical comorbidities, and have

a higher survival rate compared to their HPV-negative counterparts [4]. The standard treatment for OPC involves a course of single daily fractionated radiation with or without concurrent chemotherapy, with cisplatin being the standard of care for advanced OPC [6,7]. Both approaches incur high rates of acute and late toxicities in a young population that could experience treatment sequelae for several decades [7]. With the objective of reducing such treatment-related toxicities, recent clinical trials have aimed to de-escalate treatment for HPV-positive patients while preserving high tumor control rates [8]. Clinical trials that evaluated a substitute for cisplatin in low-risk HPV-positive patients reported inferior overall survival and progression-free survival with no reduced toxicities [9,10]. Therefore, markers beyond HPV status are essential for distinguishing between low- and high-risk subgroups and predicting favorable or unfavorable responses.

In OPC patients, the majority of progressions occur during the first 2 years after treatment [11]. Current clinical practice for treatment follow-up is based on a patient's risk of relapse. Generally, follow-up occurs every 6–12 weeks for the first year, every 3 months for the second year, and less frequently thereafter. Pre- and post-treatment 18-F-fluorodeoxyglucose positron emission tomography (PET)-computed tomography (CT) is routinely performed for OPC patients, and it plays a central role in the detection of locoregional and distant progression before patients present with symptoms. Anzi et al. reported lower mortality among patients with surveillance PET-CT imaging than patients without surveillance imaging [12]. Nevertheless, several challenges remain regarding surveillance PET-CT imaging: (i) follow-up PET-CTs can be difficult to interpret given persistent inflammatory changes that are often visualized post-treatment, (ii) post-treatment imaging remains understudied, with suboptimal schedules resulting in unnecessary scans too soon (or late) relative to when progression occurs, and (iii) the number of robust biomarkers for the early assessment of the risk of cancer progression is limited. The lack of optimal surveillance imaging paradigms and robust predictive markers of early and late local, regional, and distant progression present an unmet clinical need that must be addressed to provide guidance for personalized follow-up and treatment strategies.

Radiomics presents a new paradigm in medical imaging with a shift from visual interpretation to the extraction of quantitative descriptors, known as “features”, that cannot be reliably assessed through visual inspection [13]. Radiomics features consist of high-dimensional sets of mathematically defined quantities that can be analyzed computationally to provide diagnostic and prognostic information [14,15]. Previous studies have shown that features from pre-treatment PET images, such as standard uptake value, metabolic tumor volume, total lesion glycolysis, and textural features, provide significant prognostic information in OPC patients [16–20]. Folkert et al. investigated whether multivariable models that combine clinical and PET imaging features had better predictive power than the individual features alone [21]. In terms of local progression, the best model they reported consisted of metabolic tumor volume and a texture feature. The model retained predictive power (AUC = 0.68, sensitivity = 0.67, specificity = 0.70) when tested on an independent cohort. Vallieres et al. evaluated multivariable models using clinical, CT, and PET-CT radiomics features to predict locoregional and distant disease progression in various head-and-neck cancers [22]. For both outcomes, they demonstrated that radiomics features have increased predictive performance when combined with clinical variables. They reported a model that combined three PET-CT radiomics variables and four clinical variables as providing the best performance (accuracy = 0.67, sensitivity = 0.63, specificity = 0.68) for locoregional progression on an independent testing cohort. While several studies have built encouraging radiomics-based predictive models, the use of single-scanner and/or single-institution data, as well as the lack of stratification for HPV status, has been a key limitation for the generalizability of these predictive models.

We present a study on a cohort of HPV-positive patients who underwent pre-treatment PET-CT on various scanners and received definitive radiation with or without chemotherapy at our institution. In this context, we hypothesize that radiomics features could be used as risk-stratifying factors to guide individualized treatment and follow-up. We built

multivariable machine learning models using clinical and radiomics features to predict disease progression at two years post-treatment and performed testing on an external dataset from a different institution. In the modern era, this provides crucial prognostic information to aid in defining low- and high-risk subgroups for cancer progression after definitive radiotherapy.

2. Materials and Methods

2.1. Patient Cohorts

A retrospective chart review was performed of patients diagnosed with OPC who received definitive external beam radiotherapy (protons or x-rays) with or without chemotherapy. These patients were treated at Mayo Clinic in Rochester, Minnesota (MCR), between May 2013 and April 2020 and Mayo Clinic in Phoenix, Arizona (MCA), between August 2013 and December 2020. Inclusion criteria were laboratory-confirmed HPV-positive carcinoma and availability of pre-treatment PET-CTs and radiotherapy treatment planning CTs. Additionally, patients were included if the period between the end of treatment and the chart review was at least 2 years or if they had developed local, regional, or distant disease progression. Patients were excluded if they had previously received radiotherapy or surgery. The Institutional Review Board at Mayo Clinic approved this study.

The clinical characteristics of the patient cohorts are summarized in Table 1. The MCR and MCA cohorts were composed of 72 patients and 22 patients, respectively. The pre-treatment PET-CT scans were performed using standard clinical procedures; given the retrospective nature of this work, there were no specific technical requirements for inclusion of image acquisition protocols.

Table 1. Patient clinical and pathological data.

Radiation Type	MCR		MCA	
	X-rays	Protons	X-rays	Protons
Total Patients	47	25	15	7
Any Progression				
Yes	15	2	7	1
No	32	23	8	6
Age at Diagnosis (y)				
Mean	62.5	64.9	63.2	65.8
Range	48.1–77.7	42.3–81.0	46.4–86.9	55.1–76.3
T Category				
T0	0	0	0	0
T1	5	0	0	0
T2	15	3	4	5
T3	8	9	4	1
T4	19	13	7	1
N Category				
N0	6	1	0	1
N1	1	7	4	1
N2	35	17	9	5
N3	5	0	2	0
Concurrent Chemotherapy				
Yes	45	24	15	7
No	2	1	0	0

MCR: Mayo Clinic, Rochester, Minnesota. MCA: Mayo Clinic, Phoenix, Arizona.

The MCR dataset was used for feature engineering, model building, and model validation. The MCA group was not exposed to any feature engineering, model building, or model validation; rather, it was kept as an independent external cohort to test the performance and generalizability of the final models.

2.2. Feature Extraction

Patient demographics and clinical characteristics were collected, including age at diagnosis, primary disease site, type of radiotherapy, and type of chemotherapy. Pathological characteristics that classify tumor, node, and metastases were scored according to the 8th Edition of the American Joint Committee on Cancer (AJCC) staging system and treated as three distinct features represented only by their respective numerical descriptor. This was done since several of the data processing techniques employed are more appropriate for numerical data than categorical data. Follow-up data, including local, regional, and distant progression, as well as time to progression, were recorded. We defined the binary outcome for any type of progression (progression/progression-free) as the label for machine learning modeling. Other clinical features, such as smoking status, number of pack years, and number of chemotherapy cycles, were not included because they were not consistently reported in the patient electronic medical record. These features were removed, rather than excluding individual patients, to maintain statistical power.

MIM software (version 7.2.7, MIM software Inc., Cleveland, OH, USA) was used to develop a semi-automatic segmentation workflow based on standard uptake value (SUV) thresholding. The thresholding was employed to reduce any deviations caused by SUV variability between patients. Two volumes of interest were generated for each patient: regions of interest (ROIs) and largest ROI. For each patient, we defined the normal tissue SUV using a volumetric SUV mean of a 5 cm sphere centered in the liver. The ROI volume was delineated quantitatively using a tumor-to-normal tissue SUV ratio of 1.5. The largest ROI volume consists of the largest connected component of the ROI volume. To visually confirm that the SUV-based segmentation was consistent with the anatomical tumor location, we used rigid registration to transfer the physician-generated tumor contours from the planning CT to the pre-treatment PET image and ensured that they overlapped with the location of the ROIs and largest ROI. The volumes were resampled to $2 \times 2 \times 2 \text{ mm}^3$ to create isotropic voxels.

We extracted radiomics features using PyRadiomics, the open-source Python package, which is also involved in the efforts by the Image Biomarker Standardization Initiative team [23]. As in previous work [24,25], our in-house software incorporated Python's scikit-learn packages [26] to build a pipeline for the processing and extraction of features. For each volume of interest, we calculated 14 shape-based features and 18 intensity-based features. A total of 64 radiomics features were calculated. Because our dataset consists of imaging data from different scanners and institutions, we did not include texture-based features in our analysis. Texture-based features are highly influenced by imaging acquisition parameters and the noise profile of the scanner; therefore, heterogeneity in the source of the imaging data may introduce uncertainty on the generalizability of these second-order features. The schematic in Figure 1 illustrates the techniques sequentially employed for feature selection, dimensionality reduction, data balancing, and machine learning modeling.

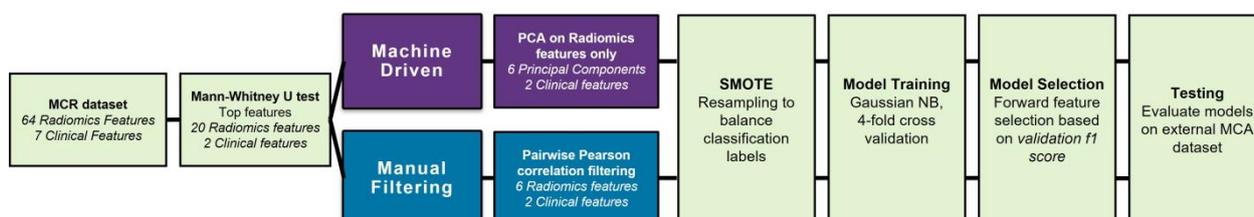


Figure 1. Schematic workflow for feature extraction, dimensionality reduction, data balancing, model training, and testing.

2.3. Dimensionality Reduction

All clinical and radiomics features were investigated for their predictive value relative to disease progression post-treatment in the MCR group. We used the Mann–Whitney U test to evaluate whether each feature was significantly different between patients with

and without progression. The features were ranked from low to high based on their p-values. The top 2 clinical features and the top 20 radiomics features were selected for further analysis. In order to further reduce the number of variables, we investigated two dimensionality reduction methods and defined a set of features for each: i) a manually filtered set and ii) a machine-driven set.

The manual filtering process consisted of computing the bivariate Pearson correlation coefficient for each possible pair of the 22 features selected in the previous section. For every group of features with a correlation coefficient larger than 0.8, one feature in each group was selected based on explainability and all other features were excluded. In other words, we selected the radiomics features with the simplest mathematical definitions that enabled understanding of the tumor's phenotype. The remaining set of 2 clinical and 6 radiomics features were normalized with a standard scaler by subtracting their mean and scaling to unit variance. The fitted scaler was saved, and the same scaler was recalled when the MCA data were tested. These 8 features were defined as the manually filtered set and used for further processing.

The machine-driven process involved the normalization and transformation of features. Each of the 22 features selected via the Mann–Whitney U test was normalized with a standard scaler by subtracting its mean and scaling to unit variance. Because a high correlation is expected for a given radiomics feature extracted from our two volumes of interest, principal component analysis (PCA) was performed on the 20 radiomics features to transform the correlated features into a smaller set of uncorrelated features (principal components) while preserving the majority of information in the dataset [26]. The generated principal components were ranked from high to low based on their explained variance, i.e., the amount of variability in the dataset explained by each component. The top 6 principal components with the largest explained variance were selected because they corresponded to the fewest number of components with an explained variance sum greater than 95%. The fitted scaler and PCA models were saved, and the same scaler and PCA models were recalled when the MCA data were tested. The machine-driven set, composed of the top 2 clinical features and the selected 6 principal components, was used for further processing.

2.4. Data Balancing

In this application, the utility of the eventual predictive model is based on its performance on the minority class, i.e., progression events. The incidence of any progression at two years post-treatment was 19.4% and 31.8% for the MCR and MCA cohorts, respectively. The imbalance between progression and progression-free events presents a challenge, in that machine learning models trained on imbalanced data may favor the class with the larger number of samples (progression-free events) and perform poorly in the prediction of progression events. To alleviate such a data imbalance problem in the MCR dataset, we adopted a combination of data over- and undersampling via SMOTETomek (The Synthetic Minority Over-Sampling Technique and Tomek), the full details of which are explained elsewhere [27]. Briefly, SMOTE creates new synthetic events by interpolating between existing progression events using k nearest neighbors [28]. Then, undersampling is performed by identifying Tomek links or pairs of events that lie closely in feature space but belong to opposite classes, and removing the event that belongs to the progression-free class [27]. The resulting dataset has the same number of progression and progression-free patients. For both the manual filtering and machine-driven methods, the implementation of SMOTETomek used the same random state seed as input. Data balancing was only performed on the MCR (training) dataset and not on the MCA (testing) dataset.

2.5. Feature Selection, Model Construction, and Testing

Several machine learning classifiers were considered due to their popularity and previous success, namely random forest, logistic regression, AdaBoost, and Gaussian Naïve Bayes (NB) [21–23,29]. Gaussian NB with default hyperparameters showed the most robust performance in terms of stability between training and cross-validation accuracy. Therefore,

in this work we present results only from the Gaussian NB classifier. A NB classifier uses the training data to learn the conditional probability of each feature given the class label. It assumes that features are mutually independent and uses Bayes' theorem to compute the probability of the classes given an instance of the features [30]. Gaussian NB assumes that the features within each class are drawn from a Gaussian distribution. For multivariable models, NB computes the product of all the probabilities from every feature and selects the class with the highest posterior probability.

Gaussian NB models were trained on the MCR dataset using each of the two sets of features, the manually filtered set and the machine-driven set. Because of the limited number of patients in the training MCR dataset and to reduce the chance of overfitting, we performed training using 4-fold cross-validation. We used a forward feature selection scheme to construct sets of models of increased complexity, i.e., number of input features from 1 to 6 features. The first step in our forward feature selection was to train all possible single-variable models and select the model with the best performance. Then, we trained incrementally more complex models by adding one feature at a time and retaining the feature that led to the best model performance. The performance metric that we used for model selection was the cross-validation f1 score, which represents the average f1 score across all 4 cross-validation trials. This approach facilitates the analysis of accumulated model learning, reveals when overfitting becomes a problem, and illustrates the complementary information provided by different features.

The 12 selected models (6 models from the manually filtered set and 6 models from the machine-driven set) were independently tested on the external MCA dataset. We evaluated prediction performance using accuracy of classification of binary endpoint (any progression at two years post-treatment).

3. Results

3.1. Dimensionality Reduction

The 2 clinical features and 20 radiomics features identified as being different between progression and progression-free patients via the Mann–Whitney U test are listed in Figure 2a in the form of a correlation matrix. The manually filtered set that resulted from correlation filtering consists of the 2 clinical features and 6 radiomics features marked with an * in Figure 2a. The 20 radiomics features listed in Figure 2a were used to generate the principal components for building the machine-driven set. As shown in Figure 2b, the fewest number of principal components with an explained variance sum greater than 95% are the first 6 principal components, which make up the machine-driven set along with the 2 clinical features. Both of these dimensionality reduction methods reduced the number of features from 71 features to 8 features in each of the two sets defined.

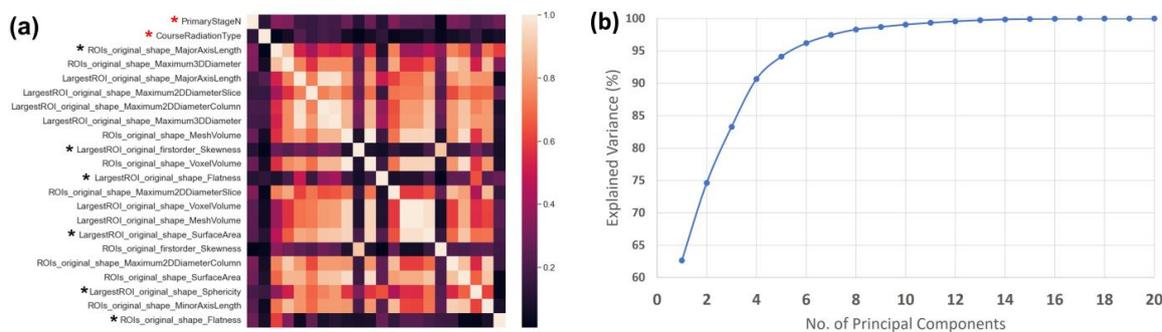


Figure 2. (a) Correlation matrix of 2 clinical and 20 radiomics features selected to be significantly different between progression and progression-free cohorts via the Mann–Whitney U test. The horizontal axis (columns) represents the same set of 22 features in the same order as presented in the vertical axis (rows). The manually filtered set is composed of 2 clinical features (red stars) and 6 radiomics features (black stars). (b) Explained variance as a function of the number of principal components obtained from PCA transformation of 20 radiomics features.

3.2. Feature Selection, Model Construction, and Testing

All possible single-variable models were trained for the manually filtered features set and machine-driven features set. The ‘Radiation Type’ clinical feature, which codes radiation modality (x-rays or protons), was identified as the best single-variable model (based on the cross-validation f1 score) in both feature sets. For each feature set, models with 1 to 6 input features were sequentially built by adding one feature at a time and retaining the feature that resulted in the best cross-validation f1 score. Table 2 summarizes the features selected by each step of our forward feature selection scheme.

Table 2. Features selected via forward feature selection scheme. Models of incremental number of features were selected by adding one feature at a time and retaining the variable that resulted in the largest cross-validation f1 score.

No. of Input Features	Dimensionality Reduction Method	
	Manually Filtered Set	Machine-Driven Set
1	Radiation Type	Radiation Type
2	Radiation Type, ROIs_MajorAxisLength	Radiation Type, PCA 3
3	Radiation Type, ROIs_MajorAxisLength, LargestROI_Flatness	Radiation Type, PCA 3, PCA 6
4	Radiation Type, ROIs_MajorAxisLength, LargestROI_Flatness, ROIs_Flatness	Radiation Type, PCA 3, PCA 6, PCA 4
5	Radiation Type, ROIs_MajorAxisLength, LargestROI_Flatness, ROIs_Flatness, LargestROI_Skewness	Radiation Type, PCA 3, PCA 6, PCA 4, Primary Stage N
6	Radiation Type, ROIs_MajorAxisLength, LargestROI_Flatness, ROIs_Flatness, LargestROI_Skewness, LargestROI_SurfaceArea	Radiation Type, PCA 3, PCA 6, PCA 4, Primary Stage N, PCA 5

PCA: Principal component analysis. Nomenclature for radiomics features: VolumeOfInterest_FeatureName.

Figure 3 shows the MCR training, MCR cross-validation, and MCA testing accuracy for each selected model. Features from the manually filtered set maintained similar cross-validation accuracy with increasing number of input features (Figure 3a); likewise, features from the machine-driven set maintained similar cross-validation accuracy for models with 2 or more features (Figure 3b), which is an indication of the robustness of using the selected features in these two sets.

The models from the manually filtered set showed improved training accuracy as the number of features increased (Figure 3a), demonstrating that the models were incrementally learning each time a new feature was added. However, the increasing difference between training accuracy and cross-validation accuracy for models with three or more features is a possible indication of overfitting. The 3-feature model for the manually filtered set had the highest cross-validation accuracy (71%); however, with a training accuracy of 73% and a testing accuracy of 54% (sensitivity = 0.86, specificity = 0.40), overfitting was confirmed, hampering the generalizability of the model.

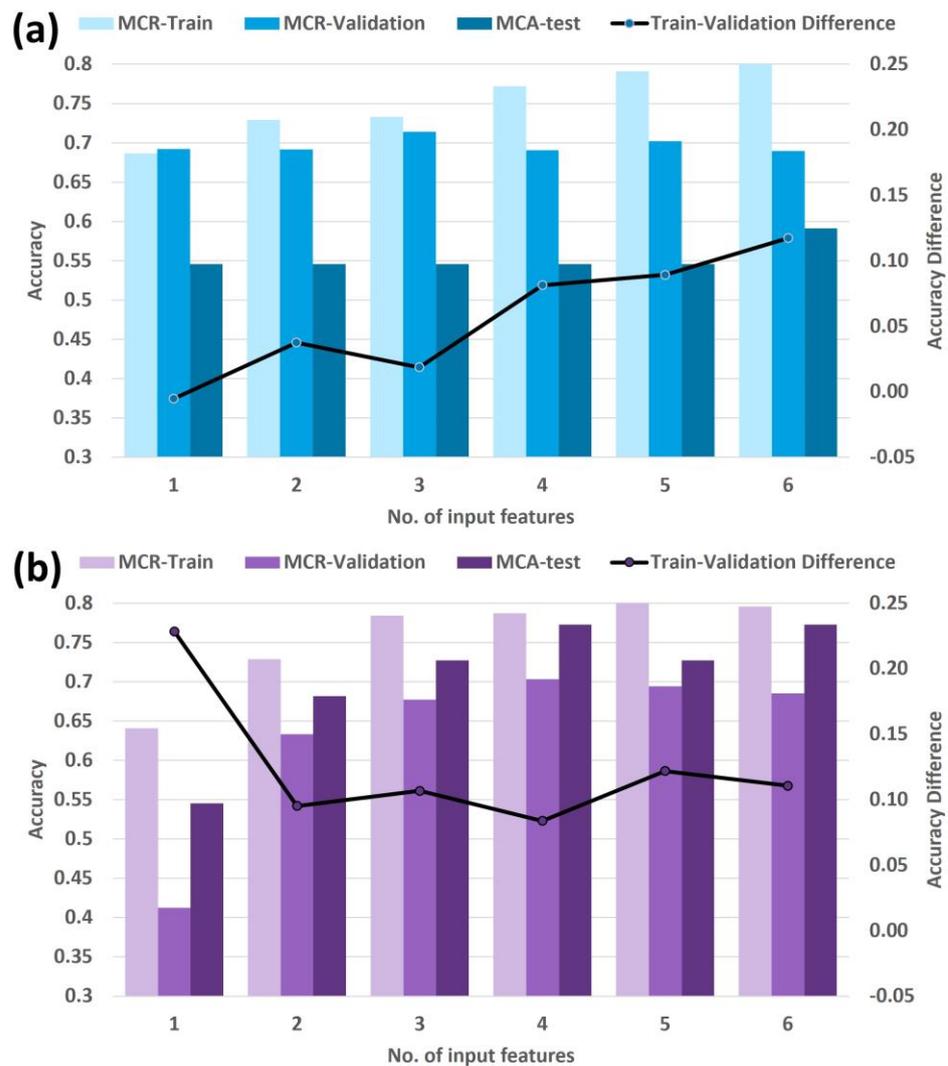


Figure 3. MCR training, MCR cross-validation, and MCA testing accuracy for models of different number of input features based on the (a) manually filtered and (b) machine-driven sets. Refer to Table 2 for the list of features used.

Similarly, the models from the machine-driven set showed incremental training accuracy up to the 5-feature model, with a slight decrease for the 6-feature model (Figure 3b). The difference between training accuracy and cross-validation accuracy decreased with the number of features with a minimum at the 4-feature model, then increasing for the 5- and 6-feature models. These trends suggest reasonable generalizability up to the 4-feature model and overfitting for the more complex models, which is supported by the rise and plateau in testing performance in Figure 3b. The 4-feature model for the machine-driven set had the highest cross-validation accuracy (70%), with training accuracy of 79% and testing accuracy of 77% (sensitivity = 0.86, specificity = 0.73).

4. Discussion

In this work, we demonstrated that a machine learning classifier based on pre-radiation clinical and PET radiomics features can be used to predict any progression of HPV-positive OPC patients at two years post-treatment. These findings are valuable for classifying patients as low- or high-risk and formulating individualized treatment and follow-up strategies after definitive radiotherapy.

This multi-institutional study consisted of a training set of MCR patients and a testing set of MCA patients. PET images were acquired following general clinical guidelines with-

out requiring specific acquisition settings for inclusion. Therefore, this work represents a multi-scanner and multi-institutional dataset, rather than a homogeneous single-institution and/or single-scanner dataset that would limit the generalizability of radiomics-based models [14].

The SUV-based normalization and thresholding method used to define the volumes of interest appears to be an effective method to reduce interpatient SUV variability and promote generalizability between scanners and institutions. While we excluded texture features from our analysis due to their instability, this behavior is consistent with previous literature. Cheng et al. indicated that textural radiomics features may be influenced by imaging processing parameters and the noise profile of the scanners [31], and Folkert et al. found that texture features obtained from two different scanners showed significant predictive power only after a smoothing method was applied [21]. Future work could consider alleviating this challenge via data harmonization to reduce the scanner effect [32].

We compared two dimensionality reduction strategies to reduce the number of clinical and radiomics features used for modeling. Our forward feature selection approach facilitated the understanding of how the incremental number of input features improved learning, and it revealed when overfitting became a problem. On the one hand, the manually selected features are simple morphological and statistical definitions describing the tumor, but they exhibited overfitting for a lower number of input features compared to the machine-driven set. Their inferior performance in multivariable models indicates a lack of complementary information between the features. On the other hand, machine-driven features generated via PCA are more difficult to interpret, but our results demonstrate that they successfully combine the complementary value of different features to inform predictions. An interesting next step would be to study these principal components and how they map to the set of manually selected features.

The use of SMOTETomek as an over- and undersampling approach proved effective in balancing the number of patients with or without progression. By creating a dataset with the same number of patients in each class, the predictions of Gaussian NB were not influenced by the prevalence of either class; rather, each prediction relied solely on the input clinical and radiomics features.

In both MCR and MCA cohorts, we noticed a lower numerical progression rate for the patients treated with proton therapy compared to x-ray therapy (MCR: 8% vs. 32%; MCA: 14% vs. 47%; Table 1). The 'Radiation Type' clinical feature was identified as the top feature from the Mann–Whitney U test. Additionally, it was detected as the best single-variable model, which suggests a strong predictive value compared to other clinical, radiomics, and principal component features. This difference cannot be explained simply by the small difference observed between patient epidemiology (see Table 1). At both institutions, the radiation dose was calculated using a relative biological effective (RBE) value of 1.1. The difference in progression (Figure 4) raises the question of the role that RBE plays in tumor control probability and whether 1.1 is an adequate value, but that inquiry is beyond the scope of this paper. Because conclusions regarding radiation type are limited by our small sample size, future studies should examine its prognostic significance in larger cohorts.

While pre- and post-treatment 18-F-fluorodeoxyglucose PET-CT is the standard for diagnosis and surveillance of head-and-neck-squamous cell cancers, others have applied different imaging modalities for radiomics-based progression prediction, including contrast-enhanced and non-contrast CT, as well as diffusion-weighted, T1, and T2 MRI sequences [14]. Future work could explore whether other imaging modalities or different PET tracers, such as fibroblast activation protein inhibitor (FAPI)-PET [33,34], offer information complementary to published 18-F-fluorodeoxyglucose PET-based models.

The presented study may be limited by selection bias and other inherent biases associated with any retrospective analysis. Additionally, the small sample size in the cohorts constitutes a limitation for the training and validation of the models. We should note that the two institutions, MCR and MCA, may conduct similar practices that could result in homogeneous treatment outcomes across the institutions. Future work should focus on

larger cohorts and additional independent cross-validation to enhance the generalizability of the presented model to multi-institutional and multi-scanner datasets.

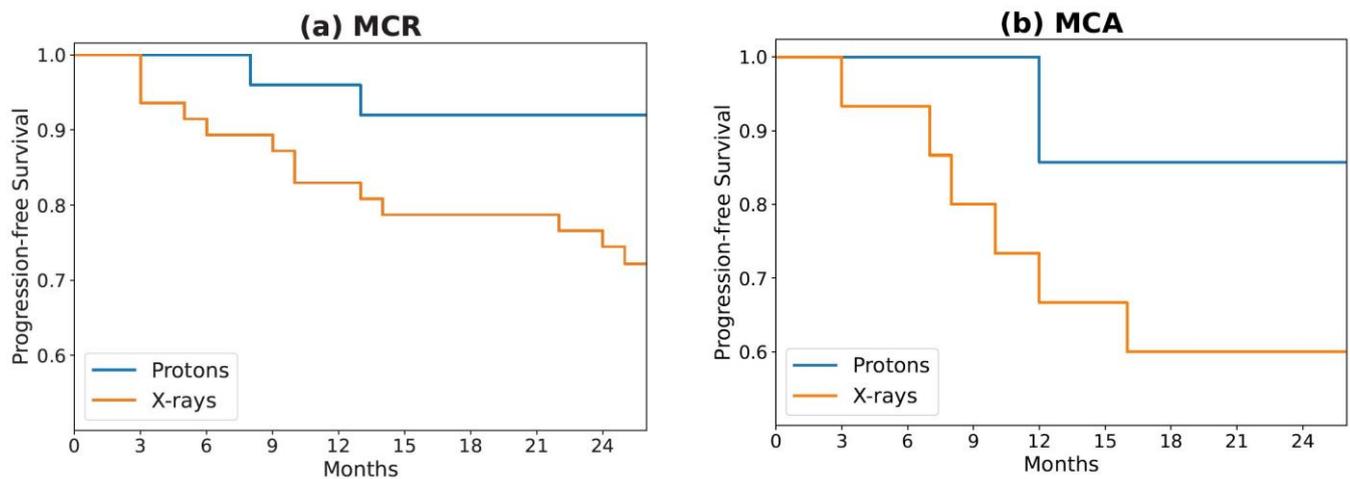


Figure 4. Kaplan–Meier survival curves for progression-free survival of (a) MCR and (b) MCA patients, stratified by the type of radiation received during treatment.

5. Conclusions

The radiomics-based classifiers built in this work used a multi-scanner dataset to predict disease progression of HPV-positive OPC patients from an external institution. Our results aid in the identification of pre-treatment clinical and radiomics features that provide risk-stratifying information for post-treatment progression. These features should be further evaluated for their ability to inform individualized treatment and follow-up strategies tailored to individual patients.

Author Contributions: Conceptualization, D.A.G., E.B.J., J.Q.; methodology, D.A.G. and J.Q.; software, D.A.G. and J.Q.; validation, D.A.G. and J.Q.; formal analysis, D.A.G. and J.Q.; investigation, D.A.G. and J.Q.; data curation, D.A.G., E.B.J., L.K.M., S.S., B.S.L., Y.R., and J.Q.; writing—original draft preparation, D.A.G.; writing—review and editing, D.A.G. and J.Q.; supervision, J.-C.M.R., R.L.F., and M.G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Mayo Clinic (protocol code 20-010245 approved on 10/13/2020).

Informed Consent Statement: Patient consent was waived due to research not involving more than a minimal risk to the subjects, and the waiver will not adversely affect the rights and welfare of the subjects.

Data Availability Statement: Data supporting the reported results can be obtained from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fakhry, C.; Krapcho, M.; Eisele, D.W.; D'Souza, G. Head and neck squamous cell cancers in the United States are rare and the risk now is higher among white individuals compared with black individuals. *Cancer* **2018**, *124*, 2125–2133. [[CrossRef](#)]
2. Chaturvedi, A.K.; Engels, E.A.; Pfeiffer, R.M.; Hernandez, B.Y.; Xiao, W.; Kim, E.; Jiang, B.; Goodman, M.T.; Sibug-Saber, M.; Cozen, W.; et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J. Clin. Oncol.* **2011**, *29*, 4294–4301. [[CrossRef](#)]
3. D'Souza, G.; Westra, W.H.; Wang, S.J.; van Zante, A.; Wentz, A.; Kluz, N.; Rettig, E.; Ryan, W.R.; Ha, P.K.; Kang, H.; et al. Differences in the Prevalence of Human Papillomavirus (HPV) in Head and Neck Squamous Cell Cancers by Sex, Race, Anatomic Tumor Site, and HPV Detection Method. *JAMA Oncol.* **2017**, *3*, 169–177. [[CrossRef](#)] [[PubMed](#)]

4. Ang, K.K.; Harris, J.; Wheeler, R.; Weber, R.; Rosenthal, D.I.; Nguyen-Tan, P.F.; Westra, W.H.; Chung, C.H.; Jordan, R.C.; Lu, C.; et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* **2010**, *363*, 24–35. [[CrossRef](#)] [[PubMed](#)]
5. Dahlstrom, K.R.; Calzada, G.; Hanby, J.D.; Garden, A.S.; Glisson, B.S.; Li, G.; Roberts, D.B.; Weber, R.S.; Sturgis, E.M. An evolution in demographics, treatment, and outcomes of oropharyngeal cancer at a major cancer center: A staging system in need of repair. *Cancer* **2013**, *119*, 81–89. [[CrossRef](#)] [[PubMed](#)]
6. Adelstein, D.J.; Li, Y.; Adams, G.L.; Wagner, H., Jr.; Kish, J.A.; Ensley, J.F.; Schuller, D.E.; Forastiere, A.A. An intergroup phase III comparison of standard radiation therapy and two schedules of concurrent chemoradiotherapy in patients with unresectable squamous cell head and neck cancer. *J. Clin. Oncol.* **2003**, *21*, 92–98. [[CrossRef](#)] [[PubMed](#)]
7. Denis, F.; Garaud, P.; Bardet, E.; Alfonsi, M.; Sire, C.; Germain, T.; Bergerot, P.; Rhein, B.; Tortochaux, J.; Calais, G. Final results of the 94-01 French Head and Neck Oncology and Radiotherapy Group randomized trial comparing radiotherapy alone with concomitant radiochemotherapy in advanced-stage oropharynx carcinoma. *J. Clin. Oncol.* **2004**, *22*, 69–76. [[CrossRef](#)]
8. Ma, D.J.; Price, K.A.; Moore, E.J.; Patel, S.H.; Hinni, M.L.; Garcia, J.J.; Graner, D.E.; Foster, N.R.; Ginos, B.; Neben-Wittich, M.; et al. Phase II Evaluation of Aggressive Dose De-Escalation for Adjuvant Chemoradiotherapy in Human Papillomavirus-Associated Oropharynx Squamous Cell Carcinoma. *J. Clin. Oncol.* **2019**, *37*, 1909–1918. [[CrossRef](#)]
9. Gillison, M.L.; Trotti, A.M.; Harris, J.; Eisbruch, A.; Harari, P.M.; Adelstein, D.J.; Jordan, R.C.K.; Zhao, W.; Sturgis, E.M.; Burtness, B.; et al. Radiotherapy plus cetuximab or cisplatin in human papillomavirus-positive oropharyngeal cancer (NRG Oncology RTOG 1016): A randomised, multicentre, non-inferiority trial. *Lancet* **2019**, *393*, 40–50. [[CrossRef](#)]
10. Mehanna, H.; Robinson, M.; Hartley, A.; Kong, A.; Foran, B.; Fulton-Lieuw, T.; Dalby, M.; Mistry, P.; Sen, M.; O’Toole, L.; et al. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): An open-label randomised controlled phase 3 trial. *Lancet* **2019**, *393*, 51–60. [[CrossRef](#)]
11. Fakhry, C.; Zhang, Q.; Nguyen-Tan, P.F.; Rosenthal, D.; El-Naggar, A.; Garden, A.S.; Soulieres, D.; Trotti, A.; Avizonis, V.; Ridge, J.A.; et al. Human Papillomavirus and Overall Survival After Progression of Oropharyngeal Squamous Cell Carcinoma. *J. Clin. Oncol.* **2014**, *32*, 3365–3373. [[CrossRef](#)]
12. Anzai, Y.; Chang, C.P.; Rowe, K.; Snyder, J.; Deshmukh, V.; Newman, M.; Fraser, A.; Smith, K.; Date, A.; Galvao, C.; et al. Surveillance Imaging with PET/CT and CT and/or MRI for Head and Neck Cancer and Mortality: A Population-based Study. *Radiology* **2023**, *307*, e212915. [[CrossRef](#)]
13. Scapicchio, C.; Gabelloni, M.; Barucci, A.; Cioni, D.; Saba, L.; Neri, E. A deep look into radiomics. *Radiol. Med.* **2021**, *126*, 1296–1311. [[CrossRef](#)]
14. Haider, S.P.; Burtness, B.; Yarbrough, W.G.; Payabvash, S. Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas. *Cancers Head Neck* **2020**, *5*, 6. [[CrossRef](#)]
15. Aerts, H.J.; Velazquez, E.R.; Leijenaar, R.T.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [[CrossRef](#)]
16. Mowery, Y.M.; Vergalaso, I.; Rushing, C.N.; Choudhury, K.R.; Niedzwiecki, D.; Wu, Q.; Yoo, D.S.; Das, S.; Wong, T.Z.; Brizel, D.M. Early (18)F-FDG-PET Response During Radiation Therapy for HPV-Related Oropharyngeal Cancer May Predict Disease Recurrence. *Int. J. Radiat. Oncol. Biol. Phys.* **2020**, *108*, 969–976. [[CrossRef](#)]
17. Kim, J.W.; Oh, J.S.; Roh, J.L.; Kim, J.S.; Choi, S.H.; Nam, S.Y.; Kim, S.Y. Prognostic significance of standardized uptake value and metabolic tumour volume on (1)(8)F-FDG PET/CT in oropharyngeal squamous cell carcinoma. *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 1353–1361. [[CrossRef](#)]
18. Lim, R.; Eaton, A.; Lee, N.Y.; Setton, J.; Ohri, N.; Rao, S.; Wong, R.; Fury, M.; Schoder, H. 18F-FDG PET/CT metabolic tumor volume and total lesion glycolysis predict outcome in oropharyngeal squamous cell carcinoma. *J. Nucl. Med.* **2012**, *53*, 1506–1513. [[CrossRef](#)]
19. Dibble, E.H.; Alvarez, A.C.; Truong, M.T.; Mercier, G.; Cook, E.F.; Subramaniam, R.M. 18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: Adding value to clinical staging. *J. Nucl. Med.* **2012**, *53*, 709–715. [[CrossRef](#)]
20. Cheng, N.M.; Fang, Y.H.; Lee, L.Y.; Chang, J.T.; Tsan, D.L.; Ng, S.H.; Wang, H.M.; Liao, C.T.; Yang, L.Y.; Hsu, C.H.; et al. Zone-size nonuniformity of 18F-FDG PET regional textural features predicts survival in patients with oropharyngeal cancer. *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 419–428. [[CrossRef](#)]
21. Folkert, M.R.; Setton, J.; Apte, A.P.; Grkovski, M.; Young, R.J.; Schoder, H.; Thorstad, W.L.; Lee, N.Y.; Deasy, J.O.; Oh, J.H. Predictive modeling of outcomes following definitive chemoradiotherapy for oropharyngeal cancer based on FDG-PET image characteristics. *Phys. Med. Biol.* **2017**, *62*, 5327–5343. [[CrossRef](#)]
22. Vallieres, M.; Kay-Rivest, E.; Perrin, L.J.; Liem, X.; Furstoss, C.; Aerts, H.; Khaouam, N.; Nguyen-Tan, P.F.; Wang, C.S.; Sultanem, K.; et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci. Rep.* **2017**, *7*, 10117. [[CrossRef](#)]
23. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [[CrossRef](#)]

24. Qian, J.; Herman, M.G.; Brinkmann, D.H.; Laack, N.N.; Kemp, B.J.; Hunt, C.H.; Lowe, V.; Pafundi, D.H. Prediction of MGMT Status for Glioblastoma Patients Using Radiomics Feature Extraction From (18)F-DOPA-PET Imaging. *Int. J. Radiat. Oncol. Biol. Phys.* **2020**, *108*, 1339–1346. [[CrossRef](#)]
25. Lucien, F.; Kim, Y.; Qian, J.; Orme, J.J.; Zhang, H.; Arafa, A.; Abraha, F.; Thapa, I.; Tryggestad, E.J.; Harmsen, W.S.; et al. Tumor-Derived Extracellular Vesicles Predict Clinical Outcomes in Oligometastatic Prostate Cancer and Suppress Antitumor Immunity. *Int. J. Radiat. Oncol. Biol. Phys.* **2022**, *114*, 725–737. [[CrossRef](#)]
26. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
27. Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]
28. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
29. Parmar, C.; Grossmann, P.; Rietveld, D.; Rietbergen, M.M.; Lambin, P.; Aerts, H.J. Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. *Front. Oncol.* **2015**, *5*, 272. [[CrossRef](#)]
30. Abraham, R.; Simha, J.B.; Iyengar, S. Medical datamining with a new algorithm for feature selection and naive bayesian classifier. In Proceedings of the 10th International Conference on Information Technology (ICIT 2007), Rourkela, India, 17–20 December 2007; pp. 44–49.
31. Cheng, N.M.; Fang, Y.H.; Chang, J.T.; Huang, C.G.; Tsan, D.L.; Ng, S.H.; Wang, H.M.; Lin, C.Y.; Liao, C.T.; Yen, T.C. Textural features of pretreatment 18F-FDG PET/CT images: Prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. *J. Nucl. Med.* **2013**, *54*, 1703–1709. [[CrossRef](#)]
32. Orlhac, F.; Eertink, J.J.; Cottreau, A.S.; Zijlstra, J.M.; Thieblemont, C.; Meignan, M.; Boellaard, R.; Buvat, I. A Guide to ComBat Harmonization of Imaging Biomarkers in Multicenter Studies. *J. Nucl. Med.* **2022**, *63*, 172–179. [[CrossRef](#)]
33. Promteangtrong, C.; Siripongsatian, D.; Jantarato, A.; Kunawudhi, A.; Kiatkittikul, P.; Yaset, S.; Boonkawin, N.; Chotipanich, C. Head-to-Head Comparison of (68)Ga-FAPI-46 and (18)F-FDG PET/CT for Evaluation of Head and Neck Squamous Cell Carcinoma: A Single-Center Exploratory Study. *J. Nucl. Med.* **2022**, *63*, 1155–1161. [[CrossRef](#)]
34. Huang, R.; Pu, Y.; Huang, S.; Yang, C.; Yang, F.; Pu, Y.; Li, J.; Chen, L.; Huang, Y. FAPI-PET/CT in Cancer Imaging: A Potential Novel Molecule of the Century. *Front. Oncol.* **2022**, *12*, 854658. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.