



Universiteit  
Leiden  
The Netherlands

## Development and external validation of a PET radiomic model for prognostication of head and neck cancer

Noortman, W.A.; Aide, N.; Vriens, D.; Arkes, L.S.; Slump, C.H.; Boellaard, R.; ... ; Geus-Oei, L.F. de

### Citation

Noortman, W. A., Aide, N., Vriens, D., Arkes, L. S., Slump, C. H., Boellaard, R., ... Geus-Oei, L. F. de. (2023). Development and external validation of a PET radiomic model for prognostication of head and neck cancer. *Cancers*, 15(10). doi:10.3390/cancers15102681

Version: Not Applicable (or Unknown)









License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/3753497>

**Note:** To cite this publication please use the final published version (if applicable).

## Article

# Development and External Validation of a PET Radiomic Model for Prognostication of Head and Neck Cancer

Wyanne A. Noortman <sup>1,2,\*</sup>, Nicolas Aide <sup>3</sup>, Dennis Vriens <sup>1</sup>, Lisa S. Arkes <sup>1,4</sup>, Cornelis H. Slump <sup>2</sup>, Ronald Boellaard <sup>5</sup>, Jelle J. Goeman <sup>6</sup>, Christophe M. Deroose <sup>7</sup>, Jean-Pascal Machiels <sup>8,9</sup>, Lisa F. Licitra <sup>10</sup>, Renaud Lhommel <sup>11</sup>, Alessandra Alessi <sup>12</sup>, Erwin Woff <sup>13</sup>, Karolien Goffin <sup>7</sup>, Christophe Le Tourneau <sup>14</sup>, Jocelyn Gal <sup>15</sup>, Stéphane Temam <sup>16</sup>, Jean-Pierre Delord <sup>17</sup>, Floris H. P. van Velden <sup>1</sup> and Lioe-Fee de Geus-Oei <sup>1,2,18</sup>

- <sup>1</sup> Section of Nuclear Medicine, Department of Radiology, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands
  - <sup>2</sup> TechMed Centre, University of Twente, 7522 NB Enschede, The Netherlands
  - <sup>3</sup> Nuclear Medicine Department, Centre Hospitalier Universitaire de Caen, 14000 Caen, France
  - <sup>4</sup> Technical Medicine, Delft University of Technology, 2628 CD Delft, The Netherlands
  - <sup>5</sup> Amsterdam University Medical Center, 1081 HV Amsterdam, The Netherlands
  - <sup>6</sup> Department of Biomedical Data Sciences, Leiden University Medical Center, 2300 RC Leiden, The Netherlands
  - <sup>7</sup> Nuclear Medicine and Molecular Imaging, Department of Imaging & Pathology, University Hospitals Leuven, KU Leuven, 3000 Leuven, Belgium
  - <sup>8</sup> Department of Medical Oncology, Institut Roi Albert II, Cliniques Universitaires Saint-Luc, 1200 Brussels, Belgium
  - <sup>9</sup> Institute for Experimental and Clinical Research (IREC, pôle MIRO), Université Catholique de Louvain (UCLouvain), 1200 Brussels, Belgium
  - <sup>10</sup> Department of Head and Neck Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milan, 20133 Milan, Italy
  - <sup>11</sup> Division of Nuclear Medicine, Institut de Recherche Clinique, Cliniques Universitaires Saint Luc, 1200 Brussels, Belgium
  - <sup>12</sup> Department of Nuclear Medicine-PET Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy
  - <sup>13</sup> Nuclear Medicine Department, Institut Jules Bordet, Université Libre de Bruxelles (ULB), Hôpital Universitaire de Bruxelles (H.U.B.), 1070 Bruxelles, Belgium
  - <sup>14</sup> Department of Drug Development and Innovation, Institut Curie, Paris-Saclay University, 75005 Paris, France
  - <sup>15</sup> Epidemiology and Biostatistics Department, Centre Antoine Lacassagne, University Côte d'Azur, 06100 Nice, France
  - <sup>16</sup> Department of Head and Neck Surgery Gustave Roussy, 94805 Villejuif, France
  - <sup>17</sup> Department of Medical Oncology, IUCT-OncoPole, 31100 Toulouse, France
  - <sup>18</sup> Department of Radiation Science & Technology, Delft University of Technology, 2628 CD Delft, The Netherlands
- \* Correspondence: w.a.noortman@lumc.nl



**Citation:** Noortman, W.A.; Aide, N.; Vriens, D.; Arkes, L.S.; Slump, C.H.; Boellaard, R.; Goeman, J.J.; Deroose, C.M.; Machiels, J.-P.; Licitra, L.F.; et al. Development and External Validation of a PET Radiomic Model for Prognostication of Head and Neck Cancer. *Cancers* **2023**, *15*, 2681. <https://doi.org/10.3390/cancers15102681>

Academic Editor: Sam Payabvash

Received: 31 January 2023

Revised: 30 April 2023

Accepted: 3 May 2023

Published: 9 May 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/>).

**Simple Summary:** A lack of external validation is still one of the major limitations of radiomics, hampering its clinical translation. The aim of this study was to build and externally validate an [<sup>18</sup>F]FDG PET radiomic model to predict overall survival in patients with head and neck squamous cell carcinoma treated with preoperative afatinib. Radiomic analysis of two cohorts of 20 and 34 patients was performed, where each cohort served once as a training and once as an external validation set. The radiomic model was compared to a clinical model and to a model that combined clinical and radiomic features. The radiomic model surpassed the clinical model in terms of predictive performance, but the combination of the radiomic and clinical model performed best. The [<sup>18</sup>F]FDG-PET radiomic signature based on the evaluation scan seems promising for the prediction of overall survival in HNSC treated with preoperative afatinib.

**Abstract:** Aim: To build and externally validate an [<sup>18</sup>F]FDG PET radiomic model to predict overall survival in patients with head and neck squamous cell carcinoma (HNSCC). Methods: Two multicentre datasets of patients with operable HNSCC treated with preoperative afatinib who underwent a baseline and evaluation [<sup>18</sup>F]FDG PET/CT scan were included (EORTC: n = 20, Unicancer: n = 34).

Tumours were delineated, and radiomic features were extracted. Each cohort served once as a training and once as an external validation set for the prediction of overall survival. Supervised feature selection was performed using variable hunting with variable importance, selecting the top two features. A Cox proportional hazards regression model using selected radiomic features and clinical characteristics was fitted on the training dataset and validated in the external validation set. Model performances are expressed by the concordance index (C-index). Results: In both models, the radiomic model surpassed the clinical model with validation C-indices of 0.69 and 0.79 vs. 0.60 and 0.67, respectively. The model that combined the radiomic features and clinical variables performed best, with validation C-indices of 0.71 and 0.82. Conclusion: Although assessed in two small but independent cohorts, an [ $^{18}\text{F}$ ]FDG-PET radiomic signature based on the evaluation scan seems promising for the prediction of overall survival for HNSCC treated with preoperative afatinib. The robustness and clinical applicability of this radiomic signature should be assessed in a larger cohort.

**Keywords:** [ $^{18}\text{F}$ ]FDG PET/CT; head and neck squamous cell carcinoma; radiomics; machine learning; overall survival; afatinib

## 1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is a common type of cancer worldwide, with approximately 900,000 cases and over 400,000 deaths in 2020 [1]. The oral cavity, oropharynx, larynx, and hypopharynx are the most frequently affected sites. Treatment options depend on disease stage and location and include surgery, radiation therapy, chemotherapy, and targeted therapy, either alone or in combination [2]. Despite advances in treatment, the prognosis for patients with advanced HNSCC remains poor, especially in advanced disease stages [3]. Recently, the tyrosine kinase inhibitor afatinib, a form of targeted therapy, has demonstrated efficacy in various types of cancer, including HNSCC [4]. In addition, its role in previously untreated HNSCC patients has been studied, where 2- [ $^{18}\text{F}$ ]fluoro-2-deoxy-D-glucose positron emission tomography ([ $^{18}\text{F}$ ]FDG PET) was used to quantify the metabolic response to afatinib therapy [5]. However, [ $^{18}\text{F}$ ]FDG PET images might contain more information than can be assessed visually or using traditional quantitative metrics. Quantitative image analysis using radiomics may provide more insights into disease biology. Radiomics is a form of medical image processing that aims to find stable and clinically relevant image-derived biomarkers for lesion characterisation, prognostic stratification, and response prediction, thereby contributing to precision medicine [6,7]. Radiomics consists of the conversion of (parts of) medical images into a high-dimensional set of quantitative features and the subsequent mining of this dataset for information useful for the quantification or monitoring of tumour or disease characteristics in clinical practice.

The field of radiomics has gained increasing attention in recent years, bringing about many proof-of-concept studies that demonstrate the feasibility and potential utility of radiomics in a wide range of applications. However, a lack of external validation is still one of the major limitations of radiomics, hampering its clinical translation. While the field of radiomics has grown and the number of papers has gradually increased over the last decade [8], widespread clinical implementation has not yet been established due to a lack of standardisation in the different steps of the radiomic pipeline and validation of the radiomic models. Individual studies show promising results, but they sometimes present spurious model interpretations caused by “cherry picking”, i.e., carrying forward only the best results obtained with a single approach while omitting contradictory results obtained with a slightly different approach [9]. Therefore, model performance should be validated using a new, independent dataset to assess whether the model is predictive for an entire target population or for a specific subset of patients only, e.g., patients with a specific demographic profile or patients who underwent a scan with a specific scanner or site-specific scan protocol. Validation using data that have neither contributed to the model design nor been kept aside from the original data set is highly recommended,

preferably using external data from different institutes [10]. External validation is of utmost importance for the clinical implementation of radiomic models.

The aim of this study was to build and externally validate an [ $^{18}\text{F}$ ]FDG PET radiomic model to predict overall survival in patients with HNSCC treated with preoperative afatinib followed by surgery.

## 2. Materials and Methods

### 2.1. Patient Population, Data Acquisition, and Image Reconstruction

[ $^{18}\text{F}$ ]FDG PET scans of two multicentric prospective cohorts (European Organisation for Research and Treatment of Cancer (EORTC) study 90,111; Unicancer Predictor) of patients with previously untreated histologically or cytologically confirmed squamous cell carcinoma of the oral cavity, oropharynx, larynx or hypopharynx were retrospectively analysed [5,11]. Patients in the treatment arm, who received preoperative afatinib (40 mg/day) for 14 to 28 days (14 days in EORTC cohort, 28 days in Unicancer cohort) and underwent [ $^{18}\text{F}$ ]FDG PET imaging on day 0 (baseline) and day 15 (evaluation) followed by surgery, were included.

In the EORTC study (ClinicalTrials.gov Identifier: NCT01538381) [5], immunohistochemical staining was performed on 4  $\mu\text{m}$  paraffin-embedded tumour sections, and p16 positivity was determined using a histology score of  $\geq 210$  according to Ang et al. (strong nuclear and/or cytoplasmic staining in  $\geq 70\%$  of tumour cells, Clone E6H4, CINtec<sup>®</sup>, Roche, Tucson, AZ, USA) [12]. PET/CT scanners were required to have EANM Research Ltd. (Vienna, Austria) (EARL) accreditation, and sites had to submit a dummy run scan following the study imaging requirements. PET acquisition was started  $66 \pm 6$  min after the injection of  $4.0 \pm 0.7$  MBq [ $^{18}\text{F}$ ]FDG per kilogram of body weight administered as an intravenous bolus. Serum glucose levels were below 7.0 mM/L. The reconstructed voxel sizes ranged from  $2.73 \times 2.73 \times 3.27$  mm<sup>3</sup> to  $4.00 \times 4.00 \times 4.00$  mm<sup>3</sup>.

In the Unicancer study (ClinicalTrials.gov Identifier: NCT01415674) [11], HPV16 was detected by single PCR with specific HPV16 primers. Other HPV subtypes were determined using multiplexed PCR on DNA extracted from frozen or formalin-fixed paraffin-embedded tumour tissue samples. Baseline and evaluation PET scans were acquired on the same scanner using similar acquisitions and reconstruction settings. PET acquisition was started  $68 \pm 11$  min after the injection of  $3.75 \pm 0.88$  MBq [ $^{18}\text{F}$ ]FDG per kilogram of body weight administered as an intravenous bolus. Serum glucose levels were below 9.0 mM/L. The reconstructed voxel sizes ranged from  $1.17 \times 1.17 \times 3.27$  mm<sup>3</sup> to  $5.47 \times 5.47 \times 5.00$  mm<sup>3</sup>.

All patients who entered the study signed an informed consent form. Both studies were conducted in accordance with the International Conference on Harmonization Good Clinical Practice standards and the Declaration of Helsinki. The EORTC trial was approved by Belgian and Italian ethics committees on 7 June 2012 and 17 July 2012, respectively, and by the Belgian and Italian health authorities on 7 May 2012 and 10 September 2012, respectively. The Predictor trial (Unicancer) was approved by a French ethics committee (Comité de Protection des Personnes Ile de France 1) on 12 July 2011 (reference 2011-mai-12618) and by the French health authorities ANSM on 21 June 2011 (reference A110602-11).

### 2.2. Quantitative Image Analysis

#### 2.2.1. VOI Delineation

Volumes of interest (VOI) delineation was performed using the Accurate tool (AmsterdamUMC, Amsterdam, The Netherlands) [13] on both the baseline and evaluation [ $^{18}\text{F}$ ]FDG PET scans using an isocontour that applies an adaptive threshold of 50% of the peak standardised uptake value ( $\text{SUV}_{\text{peak}}$ ), obtained using a sphere 12 mm in diameter [14], corrected for local background [15]. Boxing was applied to exclude surrounding [ $^{18}\text{F}$ ]FDG-avid tissues such as Waldeyer's tonsillar ring. VOIs were delineated by W.A.N. (PhD candidate in PET radiomics with 5 years of experience) and supervised by D.V. (nuclear medicine physician with 14 years of experience).

### 2.2.2. Radiomic Feature Extraction

Radiomic feature extraction was performed in PyRadiomics version 3.0 in Python version 3.7 (Python Software Foundation, Wilmington, DE, USA) [16]. For baseline as well as evaluation scans, 105 radiomic features were extracted: 18 first order features, 14 shape features, and 73 texture features (22 grey level co-occurrence matrix (GLCM), 16 grey level run length matrix, 16 grey level size zone matrix, 14 grey level dependence matrix, and 5 neighbouring grey tone difference matrix). In addition, total lesion glycolysis—the product of the mean SUV and the metabolic tumour volume—was calculated. A fixed bin size of 0.5 g/mL was applied, and images were interpolated to isotropic voxels of  $4.00 \times 4.00 \times 4.00 \text{ mm}^3$  using B-spline interpolation, with grids aligned by the input origin. In addition, delta radiomics expressing the percentual differences between baseline and evaluation features were calculated using Equation (1) [17]:

$$\Delta\text{feature} = \frac{\text{feature}_{\text{evaluation}} - \text{feature}_{\text{baseline}}}{\text{feature}_{\text{baseline}}} \times 100\% \quad (1)$$

### 2.3. Statistical Analysis

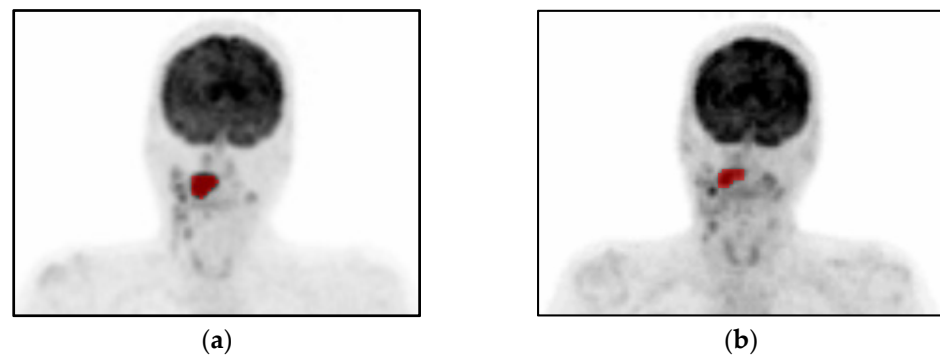
Statistical analysis was performed in R version 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria) using the packages FMRadio and randomForestSRC [18,19]. Differences in clinical characteristics between cohorts were compared using Pearson's chi-squared test for categorical variables and the Mann–Whitney U test or the independent sample t-test after testing for (log-)normality for continuous variables. Differences in survival curves were compared using log-rank statistics. As the cut-off for considering statistical significance, a type I error probability (p) of 0.05 was defined for each test.

Each cohort served once as a training set and once as an external validation set for the prediction of overall survival. Overall survival was measured from the date of inclusion to the date of death by any cause or was censored to the date of last contact. Features were standardised (centred around 0, standard deviation of 1) to prevent features with the largest scale from dominating the analysis. Redundancy filtering of the Pearson correlation matrix ( $r = 0.9$ ) was performed. Subsequently, supervised feature selection was performed using variable hunting with variable importance [20], which was repeated 1000 times, selecting the top two features (i.e., one feature per ten subjects) ranked in terms of occurrence. A Cox proportional hazards regression model using the selected radiomic features was fitted on the training dataset and validated in the external validation set. In addition, a clinical model that univariately selected the two best-performing clinical characteristics, and a combined model that combined the two radiomic features and the two clinical variables, were fitted. Model performances are expressed by the concordance index (C-index). The C-index is a generalisation of the area under the ROC curve that can take into account censored data. The proportional hazard assumption was graphically assessed by plotting the Schoenfeld residuals against time.

The findings were validated in a sham experiment [21]. The outcome labels were randomly shuffled for 100 iterations, and mean C-indices were calculated. The randomisation of the outcome labels preserves the distributions and multicollinearity of the radiomic features and the prevalence of the outcome, but it uncouples their potential relation.

## 3. Results

Twenty and thirty-four patients were included in the EORTC and Unicancer cohort, respectively (Figure 1). Three patients were excluded because no evaluation scan was available. Clinical characteristics and maximum SUV ( $\text{SUV}_{\text{max}}$ ) were not significantly different between cohorts (Table 1). The metabolic tumour volumes at baseline PET as well as at evaluation PET were significantly larger in the Unicancer cohort compared to the EORTC cohort (Table 1). Plotting the Schoenfeld residuals against time showed that all the covariates in the Cox proportional hazards model met the proportional hazard assumption (Supplementary Figure S1).



**Figure 1.** Example of [ $^{18}\text{F}$ ]FDG PET imaging in a patient with a HNSCC in the tongue (T2N2M0) with (a) the baseline scan and (b) the evaluation scan after preoperative afatinib (40 mg/day) for 14 days. Volumes of interests are indicated in red.

**Table 1.** Clinical characteristics and traditional quantitative imaging parameters of included patients.

Characteristic	EORTC (N = 20)	Umicancer (N = 34)	p-Value
Age (years), median (range)	58 (35–76)	57 (44–74)	0.911
Sex (M/F)	14/6	26/8	0.600
p16/HPV status * (positive/negative)	2/18	5/29	0.619
<b>T-stage</b>			
- T1	2	1	0.097
- T2	12	11	
- T3	1	4	
- T4	5	18	
<b>N-stage</b>			
- N0	8	14	0.932
- N+	12	20	
48-month survival (percentage $\pm$ standard deviation)	57.3 $\pm$ 13.2%	73.5 $\pm$ 7.6%	0.434 <sup>‡</sup>
SUV <sub>max</sub> baseline (g/mL), median (range)	10.6 (5.7–17.4)	12.2 (4.8–25.2)	0.257
SUV <sub>max</sub> evaluation (g/mL), median (range)	7.1 (4.2–15.4)	8.5 (3.6–24.4)	0.252
MTV baseline (cm <sup>3</sup> ), median (range)	4.9 (1.9–19.3)	7.9 (1.6–42.2)	0.032 <sup>‡</sup>
MTV evaluation (cm <sup>3</sup> ), median (range)	4.2 (1.6–9.5)	5.4 (2.0–39.2)	0.041 <sup>‡</sup>

\* p16 positivity (EORTC cohort) was considered a surrogate of HPV status (Umicancer cohort). <sup>‡</sup> Significant. <sup>‡</sup> Log-rank. SUV<sub>max</sub>: maximum standardised uptake value, MTV: metabolic tumour volume.

When the EORTC dataset served as a training set and the Umicancer dataset as an external validation set, the clinical model alone yielded a C-index of 0.60 in the validation set for the prediction of overall survival (Table 2). The selected clinical characteristics were the age of the patient and N-stage. The radiomic model performed slightly better, with a C-index of 0.69 in the validation set. Two shape features from the evaluation PET were selected: the sphericity and the maximum 2D diameter of the slice. The combined model, including both the selected clinical features and the selected radiomic features, performed similarly, with a C-index of 0.71 in the validation set.

In contrast, when the Umicancer dataset served as a training set and the EORTC dataset as an external validation set, the clinical model alone yielded a validation C-index of 0.67 (Table 3). The selected clinical characteristics were again the age of the patient and N-stage. The radiomic model performed better, with a validation C-index of 0.79, selecting two shape features from the evaluation PET: the sphericity and the maximum 2D diameter of the column. The combined model performed slightly better, with a validation C-index of 0.82.

**Table 2.** C-indices of the Cox proportional hazards model of the clinical model, the radiomic model, and the combined model. The EORTC dataset served as a training set, and the Unicancer dataset served as an external validation set.

C-Indices $\pm$ Standard Error	Training (EORTC)	Validation (Unicancer)
Clinical model: - Age - N-stage	0.70 $\pm$ 0.13	0.60 $\pm$ 0.10
Radiomic model: - Sphericity (evaluation PET) - Maximum 2D diameter slice (evaluation PET)	0.75 $\pm$ 0.11	0.69 $\pm$ 0.09
Combined model: All variables specified above	0.79 $\pm$ 0.09	0.71 $\pm$ 0.08

**Table 3.** C-indices of the Cox proportional hazards model of the clinical model, the radiomic model, and the combined model. The Unicancer dataset served as a training set, and the EORTC dataset served as an external validation set.

C-Indices $\pm$ Standard Error	Training (Unicancer)	Validation (EORTC)
Clinical model: - Age - N-stage	0.61 $\pm$ 0.10	0.67 $\pm$ 0.10
Radiomic model: - Maximum 2D diameter column (evaluation PET) - Sphericity (evaluation PET)	0.71 $\pm$ 0.08	0.79 $\pm$ 0.09
Combined model: All variables specified above	0.73 $\pm$ 0.08	0.82 $\pm$ 0.07

In the sham experiment, no model yielded a C-index different from 0.5 (range: 0.47–0.51) in the validation set.

#### 4. Discussion

In this study, we built and externally validated a radiomic model with [ $^{18}\text{F}$ ]FDG PET features to predict overall survival in patients with HNSCC treated with preoperative afatinib prior to surgery. Even though we analysed two small cohorts, model development and validation seem feasible. The radiomic models surpassed the clinical models in terms of predictive performance, but the combination of the radiomic and the clinical model showed the best performance.

HNSCC is a challenging disease to treat, particularly when it is diagnosed at an advanced stage, and early diagnosis and intervention are crucial for improving patient outcomes [3,22]. Radiomics has the potential to play an important role in the management of HNSCC by identifying patients who may benefit from a specific therapy, evaluating the risk of recurrence, and predicting survival. Several studies assessed [ $^{18}\text{F}$ ]FDG PET radiomics in HNSCC for numerous applications [23–26]. In addition, many studies have investigated radiomics extracted using different modalities to assess the effectiveness of various tyrosine kinase inhibitors in different cancer types [27–32]. However, to the best of our knowledge, no study specifically assessed [ $^{18}\text{F}$ ]FDG PET radiomics for the prognostication of HNSCC patients.

Interpretation of the radiomic features may provide insight into the semantics or tumour phenotype as captured by the PET scan. In both models, the shape feature sphericity in the evaluation PET was selected. This indicates that sphericity is a robust feature, since radiomic feature selection generally does not return identical features between methods or sometimes even between folds due to the large number of (multicollinear) features. Sphericity is a measure of the roundness of the VOI relative to a sphere. The value ranges

from 0 to 1, where a value of 1 indicates a perfect sphere. A lower sphericity, i.e., a larger surface compared to the volume, is associated with higher mortality. It is hypothesised that the sphericity of the tumour impacts its resectability: a perfectly spheric tumour might be easier to resect than an aspheric tumour with a large surface area compared to its volume. Several previous studies already described the prognostic value of primary tumour *asphericity* in (oral cavity) HNSCC determined from [<sup>18</sup>F]FDG PET [33,34] and CT [35], including cut-offs for the prediction of overall and progression-free survival. Please note that the definition of sphericity and asphericity may vary between software packages, and cut-offs might not be applicable for each software implementation. Preferably, feature definitions of the Image Biomarker Standardisation Initiative would be used [36].

In addition, both models selected a 2D diameter feature from the evaluation PET, which is defined as the largest pairwise Euclidian distance in a specific direction. The 2D diameter slice represents the diameter in the axial plane, and the 2D diameter column represents the coronal plane. This indicates that a larger tumour diameter, specifically in the transversal plane, is associated with higher mortality. Tumour size is considered an important prognostic factor in HNSCC [22,37]. Our findings are similar to those of Apostolova et al., as they present an incremental prognostic value of metabolic tumour volume (tumour dimension in 3D) in addition to asphericity [33].

It should be noted that both radiomic models only selected shape features. This might be attributed to the fact that the recommended minimal number of voxels for the texture analysis of 64 voxels per VOI was not reached in around one third of the VOIs [9]. For cubic regions, this corresponds to  $4 \times 4 \times 4 \text{ mm}^3$  voxels, i.e.,  $\sim 4 \text{ cm}^3$ . Consequently, in terms of tumour size, HNSCC are not the most suitable lesions for PET radiomic analysis. Another possible explanation for the selection of two shape features might be the heterogeneity of the datasets. We investigated two multicentre studies that included scans from four and eleven different centres, respectively. While the EORTC centres were EARL-accredited, no such standardisation and/or harmonisation was performed in the Unicancer cohort. Radiomic features are highly impacted by image acquisition and reconstruction settings [38]. A multicentre phantom study by Pfaehler et al. in three institutions using six PET/CT systems and using clinical as well as EARL1 and EARL2 reconstructions demonstrated that the percentage of intensity and texture features that yielded a moderate or better ( $>0.6$ ) intraclass correlation was low when compared to shape features ( $\sim 30\%$  and  $\sim 20\%$  vs.  $\sim 90\%$ , respectively) [39]. Moreover, they showed that EARL-compliant (preferably EARL2) reconstructions are beneficial and lead to a larger number of reliable, repeatable, and reproducible features.

Post-reconstruction harmonisation strategies such as ComBat harmonisation could have been employed to reduce heterogeneity and deal with the batch effect (i.e., the centre effect or imaging protocol effect) [40]. ComBat harmonisation is a method initially described for genomic data analysis that normalises radiomic features in order to remove the centre effect while retaining pathophysiologic information to in turn facilitate multicentre studies and exchange radiomic models between institutes. Unfortunately, with fifteen different imaging sites, the number of patients per centre was lower than the recommended twenty to thirty patients required for ComBat harmonisation [41]. ComBat harmonisation of the non-EARL-compliant imaging protocols onto the EARL-compliant imaging protocols could also be considered. However, ComBat harmonisation in prospective settings is challenging since the transformation can only be deduced from previously acquired data for the same patient population and imaging protocol [41], while in clinical practice, scanners and thereby imaging protocols are regularly replaced. In that perspective, selecting robust features irrespective of imaging protocol might be of greater interest. Since both models returned similar features, the radiomic model is considered relatively robust.

In addition, both models only selected features from the evaluation scan, disregarding the baseline and delta features. Evaluation features might be selected instead of baseline features, since evaluation features take into account the effect of the preoperative afatinib. Moreover, the delta radiomics could have been influenced by the heterogenous scanning



protocols. Although little is known about the standardisation and technical challenges of delta radiomics [42], minor differences in patient preparation, acquisition, and reconstruction settings between the baseline and evaluation PET may have introduced errors.

It is somewhat surprising that the C-indices of the EORTC validation set are higher than those of the Unicancer training set, whereas the C-index of the training set is generally higher than the C-index of the validation set. A possible explanation might be that the EORTC validation set has less complexity, which might be due to the heterogeneity in the Unicancer dataset. This explanation might be confirmed by the fact that similar features were selected in both the EORTC and Unicancer model and by the fact that the model trained by the EORTC dataset shows a higher C-index than the model trained by the Unicancer dataset.

Our study has several strengths and limitations. The main strength of our work is the external validation, which most radiomic models lack. We also showed that model development and validation is possible in small datasets in the case of a strong biological signal. A limitation of this study is that we chose not to assess low-dose CT radiomic features since we already included over 300 baseline, evaluation, and delta features in two small datasets of 20 and 34 patients, respectively. Another limitation is that clinical parameters were not always assessed in a similar way across both studies, i.e., p16 positivity was considered a surrogate for HPV status. Although they are highly correlated, subgroups of p16+/HPV− and p16−/HPV+ can be identified with different prognoses [43]. In addition, treatment response to afatinib might have been an interesting outcome measure, but this was assessed as metabolic response in one study and according to RECIST criteria in the other. Similarly, progression-free survival was only monitored in one study.

## 5. Conclusions

Although assessed in two small but independent cohorts, an [<sup>18</sup>F]FDG-PET radiomic signature based on the evaluation scan seems promising for the prediction of overall survival from HNSCC treated with preoperative afatinib. The radiomic model surpassed the clinical model in terms of predictive performance, but the combination of the radiomic and clinical model performed best. The robustness and clinical applicability of this radiomic signature should be assessed in a larger cohort.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15102681/s1>, Figure S1: Proportional hazard assumption test for overall survival by plotting the Schoenfeld residuals against time for both the EORTC and Unicancer cohort. The x-axis represents the survival time in months; the y-axis represents the scaled Schoenfeld residuals for sphericity, maximum 2D diameter slice/column, age, and N-stage.

**Author Contributions:** Conceptualization, W.A.N., D.V., C.H.S., F.H.P.v.V. and L.-F.d.G.-O.; methodology, W.A.N., D.V., F.H.P.v.V. and L.-F.d.G.-O.; software, R.B. and F.H.P.v.V.; validation, W.A.N., D.V., F.H.P.v.V. and L.-F.d.G.-O.; formal analysis, W.A.N. and L.S.A.; investigation, W.A.N. and L.S.A.; resources, N.A., C.M.D., J.-P.M., L.F.L., R.L., A.A., E.W., K.G., C.L.T., J.G., S.T. and J.-P.D.; data curation, W.A.N., D.V., F.H.P.v.V. and L.-F.d.G.-O.; writing—original draft preparation, W.A.N.; writing—review and editing, N.A., D.V., L.S.A., C.H.S., R.B., J.J.G., C.M.D., J.-P.M., L.F.L., R.L., A.A., E.W., K.G., C.L.T., J.G., S.T., J.-P.D., F.H.P.v.V. and L.-F.d.G.-O.; visualization, W.A.N.; supervision, D.V., C.H.S., J.J.G., F.H.P.v.V. and L.-F.d.G.-O.; project administration, D.V., F.H.P.v.V. and L.-F.d.G.-O.; funding acquisition, F.H.P.v.V. and L.-F.d.G.-O. All authors have read and agreed to the published version of the manuscript.

**Funding:** Boehringer Ingelheim France supported the Unicancer Predictor study with treatment supplies and financial funding.

**Institutional Review Board Statement:** Both studies were conducted in accordance with the International Conference on Harmonization Good Clinical Practice standards and the Declaration of Helsinki. The EORTC trial was approved by Belgian and Italian ethics committees on 7 June 2012 and 17 July 2012, respectively, and by the Belgian and Italian health authorities on 7 May 2012 and 10 September 2012, respectively. The Predictor trial (Unicancer) was approved by a French ethics committee (Comité de Protection des Personnes Ile de France 1) on 12 July 2011 (reference 2011-mai-12618) and by the French health authorities ANSM on 21 June 2011 (reference A110602-11).

**Informed Consent Statement:** Patients from both cohorts who entered the study signed an informed consent form.

**Data Availability Statement:** Any requests for raw and analysed data will be reviewed by a data release committee to verify whether the request is subject to any intellectual property or confidentiality obligations. Patient-related data and images in the paper were generated as part of a clinical trial and are subject to patient confidentiality. Any data and materials (for example, clinical data or imaging data) that can be shared will need approval from the EORTC and Unicancer. Any data shared will be de-identified. A signed data access agreement with the collaborator is required before accessing shared data. Requests should be made to the corresponding author (w.a.noortman@lumc.nl); response time will be within approximately 10 to 20 business days.

**Acknowledgments:** The authors want to thank Yan Liu (EORTC) and Marta Jimenez (Unicancer) for their support. We thank the patients and the families as well as the investigators who took part to the EORTC and Unicancer studies.

**Conflicts of Interest:** C.M.D. has been a consultant for Terumo, Ipsen, Sirtex, Bayer, and PSI CRO. K.G. reports fees from Bayer (consultant, speakers bureau, research funding), GE healthcare (speakers bureau), Lightpoint (consultant), PSI CRO (consultant) and Telix (consultant) outside of the submitted work. J.P.M. is an advisory board member or speaker with honoraria (managed by institution) for Pfizer, Roche, Astra/Zeneca, Bayer, Innate, Merck Serono, Boehringer, BMS, Novartis, Janssen, Incyte, Cue Biopharma, ALX Oncology, iTEOS, eTheRNA, NEKTAR, F-Star, Seagen, Genmab, Astellas; receives travel expenses from Amgen, BMS, Pfizer, MSD, Gilead; is a member of the data safety monitoring board with honoraria for Psioxus; his institution receives funding for research from all companies; he is an uncompensated advisor at MSD; and he is EORTC investigator, study coordinator, and chair of the H & N group. J.-P.D. participated in advisory boards for BMS, MSD, Pierre Fabre, and Roche; received research grants from Amgen, Astra Zeneca, BMS, Genentech, MSD, and transgene; and was an invited speaker at Merck Serono. The other authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Argiris, A.; Karamouzis, M.V.; Raben, D.; Ferris, R.L. Head and neck cancer. *Lancet* **2008**, *371*, 1695–1709. [[PubMed](#)]
3. Jakobsen, K.K.; Grønhoj, C.; Jensen, D.H.; Karnov, K.K.S.; Agander, T.K.; Specht, L.; von Buchwald, C. Increasing incidence and survival of head and neck cancers in Denmark: A nation-wide study from 1980 to 2014. *Acta Oncol.* **2018**, *57*, 1143–1151. [[CrossRef](#)] [[PubMed](#)]
4. Seiwert, T.Y.; Fayette, J.; Cupissol, D.; Del Campo, J.M.; Clement, P.M.; Hitt, R.; Degardin, M.; Zhang, W.; Blackman, A.; Ehrnrooth, E.; et al. A randomized, phase II study of afatinib versus cetuximab in metastatic or recurrent squamous cell carcinoma of the head and neck. *Ann. Oncol.* **2014**, *25*, 1813–1820. [[CrossRef](#)]
5. Machiels, J.P.; Bossi, P.; Menis, J.; Lia, M.; Fortpied, C.; Liu, Y.; Lhommel, R.; Lemort, M.; Schmitz, S.; Canevari, S.; et al. Activity and safety of afatinib in a window preoperative EORTC study in patients with squamous cell carcinoma of the head and neck (SCCHN). *Ann. Oncol.* **2018**, *29*, 985–991. [[CrossRef](#)]
6. Limkin, E.J.; Sun, R.; Dercle, L.; Zacharaki, E.I.; Robert, C.; Reuze, S.; Schernberg, A.; Paragios, N.; Deutsch, E.; Ferte, C. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann. Oncol.* **2017**, *28*, 1191–1206. [[CrossRef](#)]
7. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)]

8. Hatt, M.; Le Rest, C.C.; Tixier, F.; Badic, B.; Schick, U.; Visvikis, D. Radiomics: Data Are Also Images. *J. Nucl. Med.* **2019**, *60*, 38S–44S. [[CrossRef](#)]
9. Orlhac, F.; Nioche, C.; Klyuzhin, I.; Rahmim, A.; Buvat, I. Radiomics in PET imaging: A practical guide for newcomers. *PET Clin.* **2021**, *16*, 597–612.
10. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)]
11. Tourneau, C.L.; Delord, J.-P.; Dolivet, G.; Malard, O.; Fayette, J.; Capitain, O.; Even, C.; Hoffmann, C.; Vergez, S.; Geoffrois, L.; et al. PREDICTOR (UNICANCER GEP11): Randomized phase II study of preoperative afatinib in untreated head and neck squamous cell carcinoma (HNSCC) patients. *J. Clin. Oncol.* **2017**, *35*, 6021. [[CrossRef](#)]
12. Ang, K.K.; Harris, J.; Wheeler, R.; Weber, R.; Rosenthal, D.I.; Nguyen-Tân, 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](#)]
13. Boellaard, R. Quantitative oncology molecular analysis suite: ACCURATE. *J. Nucl. Med.* **2018**, *59*, 1753.
14. Wahl, R.L.; Jacene, H.; Kasamon, Y.; Lodge, M.A. From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors. *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.* **2009**, *50*, 122S–150S.
15. Frings, V.; van Velden, F.H.P.; Velasquez, L.M.; Hayes, W.; van de Ven, P.M.; Hoekstra, O.S.; Boellaard, R. Repeatability of Metabolically Active Tumor Volume Measurements with FDG PET/CT in Advanced Gastrointestinal Malignancies: A Multicenter Study. *Radiology* **2014**, *273*, 539–548. [[PubMed](#)]
16. 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](#)]
17. Carvalho, S.; Leijenaar, R.T.H.; Troost, E.G.C.; van Elmpt, W.; Muratet, J.P.; Denis, F.; De Ruyscher, D.; Aerts, H.; Lambin, P. Early variation of FDG-PET radiomics features in NSCLC is related to overall survival—The “delta radiomics” concept. *Radiother. Oncol.* **2016**, *118*, S20–S21. [[CrossRef](#)]
18. Ishwaran, H.; Kogalur, U.B.; Blackstone, E.H.; Lauer, M.S. Random survival forests. *Ann. Appl. Stat.* **2008**, *2*, 841–860.
19. Peeters, C.F.; Übelhör, C.; Mes, S.W.; Martens, R.; Koopman, T.; de Graaf, P.; van Velden, F.H.; Boellaard, R.; Castelijns, J.A.; Beest, D.E.T. Stable prediction with radiomics data. *arXiv* **2019**, arXiv:1903.11696.
20. Amini, M.; Hajianfar, G.; Hadadi Avval, A.; Nazari, M.; Deevband, M.R.; Oveisi, M.; Shiri, I.; Zaidi, H. Overall Survival Prognostic Modelling of Non-small Cell Lung Cancer Patients Using Positron Emission Tomography/Computed Tomography Harmonised Radiomics Features: The Quest for the Optimal Machine Learning Algorithm. *Clin. Oncol. R. Coll. Radiol.* **2022**, *34*, 114–127. [[CrossRef](#)]
21. Buvat, I.; Orlhac, F. The dark side of radiomics: On the paramount importance of publishing negative results. *J. Nucl. Med.* **2019**, *60*, 1543–1544. [[CrossRef](#)] [[PubMed](#)]
22. Wang, S.; Liu, Y.; Feng, Y.; Zhang, J.; Swinnen, J.; Li, Y.; Ni, Y. A Review on Curability of Cancers: More Efforts for Novel Therapeutic Options Are Needed. *Cancers* **2019**, *11*, 1782. [[CrossRef](#)]
23. Martens, R.M.; Koopman, T.; Noij, D.P.; Pfaehler, E.; Übelhör, C.; Sharma, S.; Vergeer, M.R.; Leemans, C.R.; Hoekstra, O.S.; Yaqub, M.; et al. Predictive value of quantitative <sup>18</sup>F-FDG-PET radiomics analysis in patients with head and neck squamous cell carcinoma. *EJNMMI Res.* **2020**, *10*, 102. [[CrossRef](#)]
24. Vallieres, M.; Kumar, A.; Sultanem, K.; El Naqa, I. FDG-PET Image-Derived Features Can Determine HPV Status in Head-and-Neck Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2013**, *87*, S467.
25. Yu, H.; Caldwell, C.; Mah, K.; Mozeg, D. Coregistered FDG PET/CT-Based Textural Characterization of Head and Neck Cancer for Radiation Treatment Planning. *IEEE Trans. Med. Imaging* **2009**, *28*, 374–383. [[CrossRef](#)]
26. Bogowicz, M.; Riesterer, O.; Stark, L.S.; Studer, G.; Unkelbach, J.; Guckenberger, M.; Tanadini-Lang, S. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. *Acta Oncol.* **2017**, *56*, 1531–1536. [[CrossRef](#)]
27. Cucchiara, F.; Del Re, M.; Valleggi, S.; Romei, C.; Petrini, I.; Lucchesi, M.; Crucitta, S.; Rofi, E.; De Liperi, A.; Chella, A.; et al. Integrating Liquid Biopsy and Radiomics to Monitor Clonal Heterogeneity of EGFR-Positive Non-Small Cell Lung Cancer. *Front. Oncol.* **2020**, *10*, 593831. [[CrossRef](#)]
28. Tang, X.; Li, Y.; Yan, W.F.; Qian, W.L.; Pang, T.; Gong, Y.L.; Yang, Z.G. Machine Learning-Based CT Radiomics Analysis for Prognostic Prediction in Metastatic Non-Small Cell Lung Cancer Patients With EGFR-T790M Mutation Receiving Third-Generation EGFR-TKI Osimertinib Treatment. *Front. Oncol.* **2021**, *11*, 719919. [[CrossRef](#)]
29. Chen, Y.-H.; Wang, T.-F.; Chu, S.-C.; Lin, C.-B.; Wang, L.-Y.; Lue, K.-H.; Liu, S.-H.; Chan, S.-C. Incorporating radiomic feature of pretreatment <sup>18</sup>F-FDG PET improves survival stratification in patients with EGFR-mutated lung adenocarcinoma. *PLoS ONE* **2021**, *15*, e0244502. [[CrossRef](#)]
30. Fan, Y.; Zhao, Z.; Wang, X.; Ai, H.; Yang, C.; Luo, Y.; Jiang, X. Radiomics for prediction of response to EGFR-TKI based on metastasis/brain parenchyma (M/BP)-interface. *Radiol. Med.* **2022**, *127*, 1342–1354. [[CrossRef](#)]

31. Zhu, J.-m.; Sun, L.; Wang, L.; Zhou, T.-C.; Yuan, Y.; Zhen, X.; Liao, Z.-W. Radiomics combined with clinical characteristics predicted the progression-free survival time in first-line targeted therapy for advanced non-small cell lung cancer with EGFR mutation. *BMC Res. Notes* **2022**, *15*, 140. [[CrossRef](#)]
32. Aerts, H.; Grossmann, P.; Tan, Y.Q.; Oxnard, G.G.; Rizvi, N.; Schwartz, L.H.; Zhao, B.S. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. *Sci. Rep.* **2016**, *6*, 33860. [[CrossRef](#)]
33. Apostolova, I.; Steffen, I.G.; Wedel, F.; Lougovski, A.; Marnitz, S.; Derlin, T.; Amthauer, H.; Buchert, R.; Hofheinz, F.; Brenner, W. Asphericity of pretherapeutic tumour FDG uptake provides independent prognostic value in head-and-neck cancer. *Eur. Radiol.* **2014**, *24*, 2077–2087.
34. Hofheinz, F.; Lougovski, A.; Zöphel, K.; Hentschel, M.; Steffen, I.G.; Apostolova, I.; Wedel, F.; Buchert, R.; Baumann, M.; Brenner, W.; et al. Increased evidence for the prognostic value of primary tumor asphericity in pretherapeutic FDG PET for risk stratification in patients with head and neck cancer. *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 429–437. [[CrossRef](#)] [[PubMed](#)]
35. Tarsitano, A.; Ricotta, F.; Cerenelli, L.; Bortolani, B.; Battaglia, S.; Lucchi, E.; Marchetti, C.; Marcelli, E. Pretreatment tumor volume and tumor sphericity as prognostic factors in patients with oral cavity squamous cell carcinoma. *J. Cranio-Maxillofac. Surg.* **2019**, *47*, 510–515. [[CrossRef](#)]
36. Zwanenburg, A.; Leger, S.; Vallieres, M.; Lock, S. Image biomarker standardisation initiative—Feature definitions v11. *arXiv* **2019**, arXiv:1612.07003.
37. Huang, S.H.; O’Sullivan, B. Overview of the 8th Edition TNM Classification for Head and Neck Cancer. *Curr. Treat. Options Oncol.* **2017**, *18*, 40. [[CrossRef](#)]
38. Traverso, A.; Wee, L.; Dekker, A.; Gillies, R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int. J. Radiat. Oncol. Biol. Phys.* **2018**, *102*, 1143–1158. [[CrossRef](#)]
39. Pfaehler, E.; van Sluis, J.; Merema, B.B.J.; van Ooijen, P.; Berendsen, R.C.M.; van Velden, F.H.; Boellaard, R. Experimental multicenter and multivendor evaluation of PET radiomic features performance using 3D printed phantom inserts. *J. Nucl. Med.* **2020**, *61*, 469–476. [[CrossRef](#)]
40. Orhac, F.; Boughdad, S.; Philippe, C.; Stalla-Bourdillon, H.; Nioche, C.; Champion, L.; Soussan, M.; Frouin, F.; Frouin, V.; Buvat, I. A Postreconstruction Harmonization Method for Multicenter Radiomic Studies in PET. *J. Nucl. Med.* **2018**, *59*, 1321–1328. [[CrossRef](#)]
41. Orhac, 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](#)] [[PubMed](#)]
42. Nardone, V.; Reginelli, A.; Grassi, R.; Boldrini, L.; Vacca, G.; D’Ippolito, E.; Annunziata, S.; Farchione, A.; Belfiore, M.P.; Desideri, I.; et al. Delta radiomics: A systematic review. *Radiol. Med.* **2021**, *126*, 1571–1583. [[CrossRef](#)] [[PubMed](#)]
43. Albers, A.E.; Qian, X.; Kaufmann, A.M.; Coordes, A. Meta analysis: HPV and p16 pattern determines survival in patients with HNSCC and identifies potential new biologic subtype. *Sci. Rep.* **2017**, *7*, 16715. [[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.