Article

Combination of Radiomics Features and Functional Radiosensitivity Enhances Prediction of Acute Pulmonary Toxicity in a Prospective Validation Cohort of Patients with a Locally Advanced Lung Cancer Treated with VMAT-Radiotherapy

Vincent Bourbonne [1,2,*] , François Lucia [1,2], Vincent Jaouen [2,3], Olivier Pradier [1,2], Dimitris Visvikis [2] and Ulrike Schick [1,2]

1 Radiation Oncology Department, University Hospital, 29200 Brest, France
2 LaTIM, INSERM, UMR 1101, University of Western Brittany, 29200 Brest, France
3 Institut Mines-Télécom Atlantique, 29200 Brest, France
* Correspondence: vincent.bourbonne@chu-brest.fr; Tel.: +33-298223398; Fax: +33-298223087

Simple Summary: Despite strict application of dose constraints, acute pulmonary toxicity (APT) remains frequent, and may impact treatment compliance and patient quality of life. Prediction models based on either a radiomics approach or a voxel-based approach were previously developed, but never combined. Combination of radiomics features and functional radiosensitivity enhances prediction of acute pulmonary toxicity. Correction of imbalanced data enhances prediction of APT. Use of such prediction models opens the possibility of tailored dosimetry planning.

Abstract: Introduction: The standard of care for people with locally advanced lung cancer (LALC) who cannot be operated on is (chemo)-radiation. Despite the application of dose constraints, acute pulmonary toxicity (APT) still often occurs. Prediction of APT is of paramount importance for the development of innovative therapeutic combinations. The two models were previously individually created. With success, the Rad-model incorporated six radiomics functions. After additional validation in prospective cohorts, a Pmap-model was created by identifying a specific region of the right posterior lung and incorporating several clinical and dosimetric parameters. To create and test a novel model to forecast the risk of APT in two cohorts receiving volumetric arctherapy radiotherapy (VMAT), we aimed to include all the variables in this study. Methods: In the training cohort, we retrospectively included all patients treated by VMAT for LALC at one institution between 2015 and 2018. APT was assessed according to the CTCAE v4.0 scale. Usual clinical and dosimetric features, as well as the mean dose to the pre-defined Pmap zone (DMean$_{Pmap}$), were processed using a neural network approach and subsequently validated on an observational prospective cohort. The model was evaluated using the area under the curve (AUC) and balanced accuracy (Bacc). Results: 165 and 42 patients were enrolled in the training and test cohorts, with APT rates of 22.4 and 19.1%, respectively. The AUCs for the Rad and Pmap models in the validation cohort were 0.83 and 0.81, respectively, whereas the AUC for the combined model (Comb-model) was 0.90. The Bacc for the Rad, Pmap, and Comb models in the validation cohort were respectively 78.7, 82.4, and 89.7%. Conclusion: The accuracy of prediction models were increased by combining radiomics, DMean$_{Pmap}$, and common clinical and dosimetric features. The use of this model may improve the evaluation of APT risk and provide access to novel therapeutic alternatives, such as dose escalation or creative therapy combinations.

Keywords: radiation pneumonitis; lung cancer; prediction; cluster of voxel; radiomics; personalized medicine

1. Introduction

In inoperable patients, radiotherapy (RT) combined with chemotherapy (Ch) is the standard of care in stage III lung cancer [1], according to international guidelines [2]. Modern radiation techniques, such as intensity-modulated radiotherapy (IMRT) and volumetric-arctherapy (VMAT), enable a higher conformation, resulting in lower doses to organs at risk (OARs) without compromising coverage of the planning target volume (PTV). Despite these advances, acute pulmonary toxicity (APT) remains frequent with a 5–25% rate [3].

APT appears to be significantly correlated to higher doses delivered to the lungs and possibly to the heart. However, strict application of organ-based dose constraints appear insufficient for prevention of such toxicities. This may be explained by the fact that dose volume histograms (DVH) do not efficiently account for spatial dose distribution or organ architecture. Radiomic features could better apprehend dose distribution heterogeneity [4] and, when applied to dose maps, appear as efficient predictors of APT [5]. A second explanation could be the lack of functional organ-based dose constraints. For this purpose, several techniques were developed for the identification of such functional sub-regions [6–9]. Most of these techniques require performing costly nuclear imaging, such as ^{68}Ga PET-CT or Ventilation-Perfusion SPECT-CT, and raise multiple questions regarding inter-imaging modality registration and guidelines for the definition of functional volumes. Based on retrospective data, a voxel-based analysis approach selected voxels significantly associated with the APT endpoint, defining a significance map [10–12]. In patients treated with VMAT-radiotherapy for lung cancer, a volume was identified in the posterior right lung. Higher doses to this volume (Pmap) were associated with a significantly higher risk of APT. Despite robust results, this model suffered from overfitting with loss of performance in the testing cohort. The radiomics model was only internally validated, and prospective evaluation remains to be tested. Furthermore, the possibility of the combination between the radiomics and Pmap approaches was never tested, nor was its interest evaluated. Given the relatively low rate of grade ≥2 APT (20–25%), the combination of the two approaches could lead to a higher performance, but only in the training cohort, due to overfitting.

In this study, we aim to combine the radiomics and voxel-based approaches and develop a new model that enhances the prediction of APT in patients treated for a locally advanced lung cancer treated with a VMAT-based RT.

2. Materials and Methods

2.1. Population

A previously presented population was used, consisting of 165 retrospectively and 42 prospectively included patients [12]. All patients were treated with curative intent for a histologically-proven locally advanced lung cancer (non-small cell or small-cell lung cancer) between 2015 and 2020. Patients with aged <18 years old and follow-up less than 1 year after RT completion were excluded. Similarly, patients with a history of thoracic radiation therapy, pneumonectomy, or incomplete RT were excluded. When performed, Ch could be delivered as a sequential or concomitant treatment. RT was delivered with a prescription dose of 60 to 66 Gy to the PTV, with 95% of the dose covering 95% of the prescription volume. All RT courses were delivered with VMAT. The study was approved by the hospital ethical committee (Retrospective cohort: NCT04545658, Prospective cohort: NCT03931356).

2.2. Toxicities

Using the CTCAE v4.0, acute pulmonary toxicity was graded as follows: APT3 was defined as a binary outcome relating to a grade 3 pulmonary toxicity event happening during the first 6 months after the start of RT, whereas APT2 was defined as a binary outcome related to a grade 2 pulmonary toxicity event occurring during the first 6 months after the start of RT. APT was retrospectively gathered for every patient in the retrospective cohort. APT was measured and prospectively collected as a secondary outcome for the prospective cohort. Actuarial incidences were employed.

2.3. Clinical and Dosimetric Features

Age, gender, performance status (PS), administration of Ch, tumor stage as determined by the American Joint Committee on Cancer (AJCC), mean expiratory volume/second (MEVS), history of smoking, and chronic obstructive pulmonary disease were all gathered as standard clinical features (COPD).

Regarding dosimetric features, all usual dose constraints for the homolateral lung (LungH), contralateral lung (LungC), both lungs (Lungs) and the heart were considered. The full list of dosimetric features is available as Supplementary Table S1.

2.4. Radiomics and Pmap Features

When combined in the radiomics-based model (Rad-model), six radiomic features were previously identified as highly correlated with the risk of APT_2 [5]: the Variance (LungH_Variance), Difference Variance (LungH_DVAR), Contrast (LungH_Contrast), and Entropy (LungH_Entropy), extracted from the LungH volume based on the Co-occurence Matrix, as well as the IC1 Information measure of correlation (LungH_IC1) extracted from the LungH volume, and the Energy extracted from the lung volume and based on the histogram (Lungs_Energy). All radiomics features were extracted using the in-house IBSI-compliant software [13] (MIRAS©), with a fixed bin number of 1 (1 Gy = 1 grey-level).

A region (Pmap) was previously identified in the posterior right lung [12]. Combining the mean dose to this Pmap region ($DMean_{Pmap}$) with 10 other features, a Pmap-model was built, strongly associated with the risk of $APT \geq$ grade 2. All features used for the Pmap-model [12] were thus considered for the model building and optimization.

2.5. Statistical Analysis

The overall population was considered as two separate and independent cohorts: the training cohort, regrouping the 165 retrospectively evaluated patients, and the testing cohort, consisting of the 42 prospectively included patients, as previously presented [12]. Feature set selection and model building were developed on the training cohort, blinded from the testing set. Each model was then evaluated on the testing cohort. Several prediction models were separately built using a neural network (NN) approach. Based on this NN approach, the radiomics and Pmap models were previously presented using respectively 6 [5] and 11 [12] features. Given the fact that the Rad-model was developed on a subset of the actual training cohort, the Rad-model was retrained. The Pmap-model was tested as previously developed.

All clinical, dosimetric, and radiomics features were considered for the development of the Combined model (Comb-model), using the same NN approach. The Multilayer Perceptron is a tool embedded in SPSS Modeler v18.0, producing a prediction model based on the APT endpoint and ranking each feature by its importance. To reduce the number of included features, features were combined using a decremental approach; at each step, the least important feature is deleted, and the model is retrained until only one feature remains. For every combination, the model was stabilized using a bootstrap approach with n = 1000 replications. For each set of features, only the model with the highest mean accuracy based on the 1000 replications was retained. A probability threshold was defined as the value maximizing the Youden Index defined by

$$YI = Sensitivity\ (Sp) + Specificity\ (Se) - 1.$$

Given the low rate of APT_2 and risk of unbalanced data, the impact of a correction using oversampling via the Synthetic Minority Oversampling Technique (SMOTE) package was evaluated [14]. Six prediction models were thus developed:

- Three without SMOTE: $Rad_{NonSmote}$, $Pmap_{NonSmote}$, and $Comb_{NonSmote}$
- Three with SMOTE: Rad_{Smote}, $Pmap_{Smote}$, and $Comb_{Smote}$

The performance of each model was then analyzed using sensitivity (Se), specificity (Sp), balanced accuracy (Bacc: mean of the Se and Sp), F1-score, and positive (PPV) and negative

(NPV) prediction values in both the training and testing cohorts. Decision curves analysis (DCA) for the training and testing sets was performed plotting the net benefit with its corresponding threshold. Precision-recall curves were also used for model comparisons.

Given the clinical importance of Grade 3 APT (APT_3), a secondary analysis was performed for the prediction of APT_3. New cut-offs were defined on the training cohort, maximizing the Youden Index as presented, but for the prediction of APT_3.

In the event of radiomic features being retained in the Combined models, an analysis of its regional significance was planned. For practicality, only the most important radiomic feature retained in the $Comb_{Smote}$ model was analyzed. Three-dimensional radiomics maps were calculated for each patient based on the lung volume, using the Pyradiomics toolbox [15], with a kernel set to 3. We then used the same methodology as previously used for the definition of the Pmap-region. The protocol was previously fully detailed [10,12]. Briefly, radiomics maps were registered to a thoracic phantom using a segment-based registration. A voxel-based analysis was then performed, with n = 10.000 replications, using a *t*-test approach evaluating the significance of each voxel for differentiating patients with APT_2 from the patients without APT_2. Correction for multiple testing was used (family-wise error). The significance map based on the most important radiomic feature retained in the $Comb_{Smote}$ model was named: PmapRad. Overlap between the Pmap and PmapRad maps was evaluated using the DICE coefficient.

3. Results

3.1. Population

The population characteristics have been previously presented [12,16]. For completeness, patient characteristics according to the occurrence of APT_2 in each cohort are summarized in Supplementary Table S2. The rate of APT $\geq$ grade 2 was 22.4 and 19.1% in the training and testing cohorts, respectively, whereas only 11 patients presented with an APT $\geq$ grade 3 (9 in the training and 2 in the testing sets, respectively).

3.2. Radiomics and Pmap-Models

The retrained $Rad_{NonSmote}$ and Rad_{Smote} models achieved AUCs of 0.91/0.85 in the training cohort (Table 1) and 0.83/0.83 in the testing cohort (Table 2). Performance of the $Rad_{NonSmote}$ and Rad_{Smote} models, according to the number of retained features, is available as Figure 1a,b, respectively, with the chosen model combining the six radiomic features. Ranking of the six included radiomics features was different between the two radiomics models. The radiomics models were obtained combining the six radiomic features, the most important feature being the $Energy\_Hist_{Lungs}$, accounting for 58.5% of the NN-model (Table 3), whereas $Variance\_Cooc_{LungH}$ ranked as the most important (43.0%) for the Rad_{Smote} model. In the testing cohort, Baccs were 82.0/78.7% with 18/24.0% probability thresholds for the $Rad_{NonSmote}$ and Rad_{Smote}, respectively (Table 2).

The $Pmap_{NonSmote}$ model was previously presented [12] and resulted in an AUC of 1.00 and 0.81 in the training and testing sets, respectively. Consequently, the deducted Baccs were 99.2% in the training set (Table 1) and 82.0% in the testing set (Table 2). Adding a correction of unbalanced data using the SMOTE approach, the $Pmap_{Smote}$ model combined eight features and resulted in a Bacc of 82.0% in the testing cohort (6% threshold). Performance of the $Pmap_{NonSmote}$ and $Pmap_{Smote}$ models, according to the number of retained features, is available as Figure 1c,d, respectively.

As planned, 3D radiomics maps were extracted for each patient, using the LungH_Variance feature. Overlap between the PmapRad and Pmap maps is presented as Supplementary Figure S1 and appears to be relatively low (DICE coefficient of 0.35).
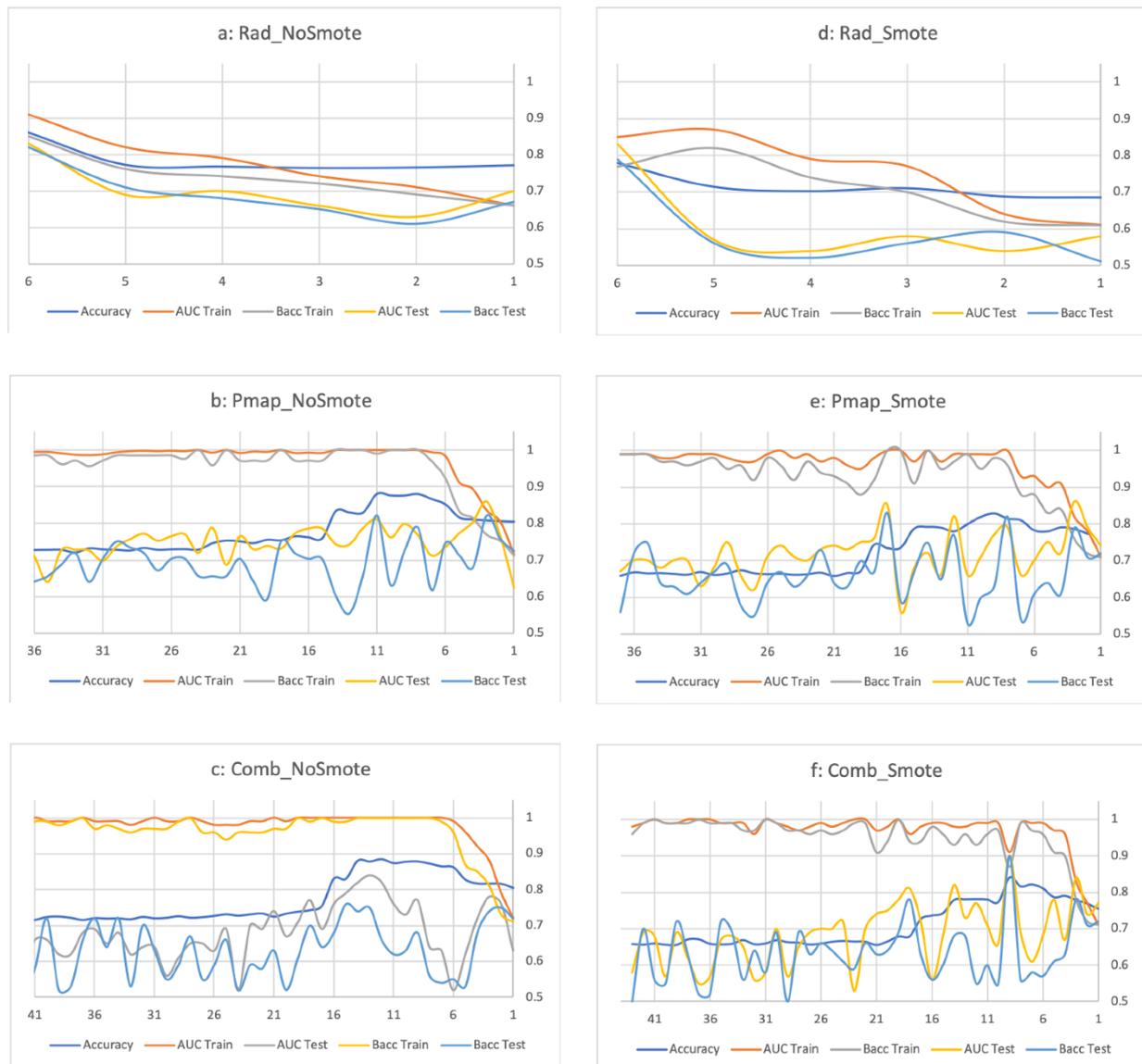
Figure 1. Performance of each model (Rad$_{NoSmote}$ (**a**), Rad$_{Smote}$ (**b**), Pmap$_{NoSmote}$ (**c**), Pmap$_{Smot}$ (**d**), Comb$_{NoSmote}$ (**e**), and Comb$_{Smote}$ (**f**)) according to the number of retained features for the prediction of APT ≥ grade 2. Abbreviation: Accuracy: Mean accuracy based on the bootstrap n = 1000 replications, APT: Acute pulmonary toxicity, Rad_NoSmote: Radiomics model without Smote, Rad_Smote: Radiomics model with Smote, Pmap_NoSmote: Pmap model without Smote, Pmap_Smote: Pmap model with Smote, Comb_NoSmote: Combined model without Smote, Comb_Smote: Combined model with Smote. Legend: x-axis: number of features in the model, y-axis: mean accuracy (n = 1000 bootstrap replications) of the model.

Table 1. Analysis of each model's discrimination between patients with or without APT $\geq$ grade 2 in the training set.

| Set | AUC | p | Cut-Off | C-Index | Se | Sp | BAcc | F1 | Number of Patients, n (%) | | | | | |
| | | | | | | | | | Below the Cutoff (Low Risk of APT $\geq$ Grade 2) | | | Above the Cutoff (High Risk of APT $\geq$ Grade 2) | | |
									Total	Without APT	With APT	Total	Without APT	With APT
Rad$_{NoSmote}$	0.91	<0.0001	>18%	0.85	86.5	84.4	85.5	0.72	113 (68.5%)	108 (95.6%)	5 (4.4%)	52 (31.5%)	20 (38.5%)	32 (61.5%)
Rad$_{Smote}$	0.85	<0.0001	>24%	0.77	75.7	78.1	76.9	0.66	109 (66.1%)	100 (91.7%)	9 (8.3%)	56 (33.9%)	28 (50.0%)	28 (50.0%)
Pmap$_{NoSmote}$	0.99	<0.0001	>8%	0.99	100.0	98.4	99.2	1.00	126 (76.4%)	126 (100.0%)	0 (0.0%)	39 (23.6%)	2 (5.1%)	37 (94.9%)
Pmap$_{Smote}$	0.99	<0.0001	>6%	0.96	100.0	91.4	95.7	0.96	117 (70.9%)	117 (100.0%)	0 (0.0%)	48 (29.1%)	11 (22.9%)	37 (77.1%)
Comb$_{NoSmote}$	0.99	<0.0001	>8%	0.99	100	98.4	99.2	1.00	126 (76.4%)	126 (100.0%)	0 (0.0%)	39 (23.6%)	2 (5.1%)	37 (94.9%)
Comb$_{Smote}$	0.91	<0.0001	>12%	0.87	81.1	92.2	86.7	0.85	125 (75.8%)	118 (94.4%)	7 (5.6%)	40 (24.2%)	10 (25.0%)	30 (75.0%)

Abbreviations: AUC: Area under the curve, Se: Sensitivity, Sp: Specificity, BAcc: Balanced Accuracy, F1: F1-score, APT: Acute pulmonary toxicity, Rad_NoSmote: Radiomics model without Smote, Rad_Smote: Radiomics model with Smote, Pmap_NoSmote: Pmap model without Smote, Pmap_Smote: Pmap model with Smote, Comb_NoSmote: Combined model without Smote, Comb_Smote: Combined model with Smote.

Table 2. Analysis of each model's discrimination between patients with or without APT ≥ grade 2 in the testing set.

Set	AUC	*p*	Cut-Off	C-Index	Se	Sp	BAcc	F1	Number of Patients, n (%)					
									Below the Cutoff (Low Risk of APT ≥ Grade 2)			Above the Cutoff (High Risk of APT ≥ Grade 2)		
									Total	Without APT	With APT	Total	Without APT	With APT
Rad$_{NoSmote}$	0.83	<0.0001	>18%	0.82	87.5	76.5	82.0	0.63	27 (64.3%)	26 (96.3%)	1 (3.7%)	15 (35.7%)	8 (53.3%)	7 (46.7%)
Rad$_{Smote}$	0.83	<0.0001	>24%	0.79	75.0	82.4	78.7	0.60	30 (71.4%)	28 (93.3%)	2 (6.7%)	12 (28.6%)	6 (50.0%)	6 (50.0%)
Pmap$_{NoSmote}$	0.81	<0.0001	>8%	0.82	87.5	76.5	82.0	0.61	27 (64.3%)	26 (96.3%)	1 (3.7%)	15 (35.7%)	8 (53.3%)	7 (46.7%)
Pmap$_{Smote}$	0.79	<0.0001	>6%	0.82	100.0	64.7	82.4	0.57	22 (52.4%)	22 (100.0%)	0 (0.0%)	20 (47.6%)	12 (60.0%)	8 (40.0%)
Comb$_{NoSmote}$	0.83	<0.0001	> 8%	0.70	62.5	76.5	69.5	0.57	29 (69.0%)	26 (89.7%)	3 (10.3%)	13 (31.0%)	8 (61.5%)	5 (38.5%)
Comb$_{Smote}$	0.90	<0.0001	>12%	0.90	100.0	79.4	89.7	0.71	27 (64.3%)	27 (100.0%)	0 (0.0%)	15 (35.7%)	7 (46.7%)	8 (53.3%)

Abbreviations: AUC: Area under the curve, Se: Sensitivity, Sp: Specificity, BAcc: Balanced accuracy, F1: F1 score, APT: Acute pulmonary toxicity, Rad_NoSmote: Radiomics model without Smote, Rad_Smote: Radiomics model with Smote, Pmap_NoSmote: Pmap model without Smote, Pmap_Smote: Pmap model with Smote, Comb_NoSmote: Combined model without Smote, Comb_Smote: Combined model with Smote.

Table 3. Classification of each feature by importance for each model.

Rad$_{NoSmote}$ Model		Rad$_{Smote}$ Model		Pmap$_{NoSmote}$ Model		Pmap$_{Smote}$ Model		Comb$_{NoSmote}$ Model		Comb$_{Smote}$ Model	
Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance
LungH_Variance	0.1%	LungH_Entropy	1.9%	V40$_{Heart}$	5.0%	Stage	3.6%	V40$_{Heart}$	4.4%	V13$_{LungH}$	3.3%
LungH_DVAR	1.5%	Lungs_Energy	4.5%	DMean$_{LungH}$	5.0%	V5$_{LungH}$	5.0%	DMean$_{LungH}$	5.0%	V5$_{LungH}$	3.8%
LungH_Contrast	3.2%	LungH_IC1	15.6%	V5$_{LungH}$	5.0%	MEVS	5.4%	V5$_{LungH}$	5.0%	V10$_{LungH}$	4.4%
LungH_IC1	15.2%	LungH_DVAR	16.9%	AJCC Stage	5.0%	V10$_{LungH}$	6.1%	AJCC Stage	5.1%	DMean$_{LungH}$	4.8%
LungH_Entropy	20.5%	LungH_Contrast	18.1%	V10$_{LungH}$	6.0%	DMean$_{LungH}$	6.6%	V10$_{LungH}$	5.5%	V30$_{2Lungs}$	6.3%
Lungs_Energy	58.5%	LungH_Variance	43.0%	COPD	6.0%	V30$_{2Lungs}$	10.2%	COPD	6.1%	COPD	10.4%
				MEVS	7.0%	DMean$_{2Lungs}$	15.5%	MEVS	6.1%	DMean$_{2Lungs}$	10.5%
				Smoking Status	7.0%	DMean$_{Pmap}$	47.7%	Smoking Status	6.2%	LungH_Variance	12.6%
				V30$_{2Lungs}$	7.0%			V30$_{2Lungs}$	6.7%	DMean$_{Pmap}$	44.1%
				DMean$_{2Lungs}$	11.0%			LungH_Variance	7.1%		
				DMean$_{Pmap}$	36.0%			DMean$_{2Lungs}$	10.1%		
								DMean$_{Pmap}$	32.7%		

Abbreviations: V40$_{Heart}$: Volume of the heart receiving 40 Gy, DMean$_{LungH}$: Mean dose received by the homolateral lung, V5$_{LungH}$: Volume of the homolateral lung receiving 5 Gy, AJCC Stage: American Joint Committee on Cancer, V10$_{LungH}$: Volume of the homolateral lung receiving 10 Gy, COPD: Chronic obstructive pulmonary disease, MEVS: Mean expiratory volume/second, V30$_{2Lungs}$: Volume of the 2 lungs receiving 30 Gy, DMean$_{2Lungs}$: Mean dose received by the 2 lungs, DMean$_{Pmap}$: Mean dose received by the Pmap volume, LungH: Homolateral lung, Lungs: both lungs, LungH_Variance: Variance extracted from the LungH volume on the Co-occurence matrix (Cooc), LungH_DVAR: Difference variance extracted from the LungH volume on the Cooc matrix, LungH_Contrast: Contrast extracted from the LungH volume on the Cooc matrix, LungH_IC1: Information measure of correlation extracted from the LungH volume on the Cooc matrix, LungH_Entropy: Entropy extracted from the LungH volume on the Cooc matrix, LungH, Lungs_Energy: Entropy extracted from the LungH volume on the histogram. Abbreviation: Rad_NoSmote: Radiomics model without Smote, Rad_Smote: Radiomics model with Smote, Pmap_NoSmote: Pmap model without Smote, Pmap_Smote: Pmap model with Smote, Comb_NoSmote: Combined model without Smote, Comb_Smote: Combined model with Smote.

3.3. Combined Model

For the combined models, 43 features were considered and their correlation with the risk of APT_2 was evaluated (Supplementary Table S3). Without SMOTE, the best model achieved a mean accuracy of 0.88 (Figure 1e) and combined 12 different features (Table 3). In this model ($Comb_{NonSmote}$), the three most important features were the $DMean_{Pmap}$, $Dmean_{2Lungs}$, and $Variance\_Cooc_{LungH}$, with respective importances of 32.7, 10.1, and 7.1%. On the training cohort, the $Comb_{NonSmote}$ model achieved an AUC of 0.91 and a Bacc of 86.2% when applying a 8% threshold (Table 1). On the testing cohort, this model resulted in a Bacc of 69.5%. Correct classification occurred for 89.7% of low-risk and 38.5% of high-risk patients (Table 2).

Use of the SMOTE approach resulted in the development of a different combined model (Figure 1f), the $Comb_{Smote}$ model, combining nine features, with the three most important features remaining the same: $DMean_{Pmap}$, $Variance\_Cooc_{LungH}$, and $Dmean_{2Lungs}$, with respective importances of 44.1, 12.6, and 10.5% (Table 3). In the testing cohort, the $Comb_{Smote}$ model resulted in a Bacc of 89.7% using the 12% threshold, with the highest PPV (53.3%) while maintaining a 100% NPV.

3.4. Model Comparison for the Prediction of APT ≥ Grade 2 and APT ≥ Grade 3

For APT_2, the $Pmap_{NoSmote}$, $Pmap_{Smote}$, and $Comb_{NoSmote}$ achieved the highest results based on the curves in the training cohort (Figure 2a), whereas no significant differences were observed between the six models in the testing cohort (Figure 2b).
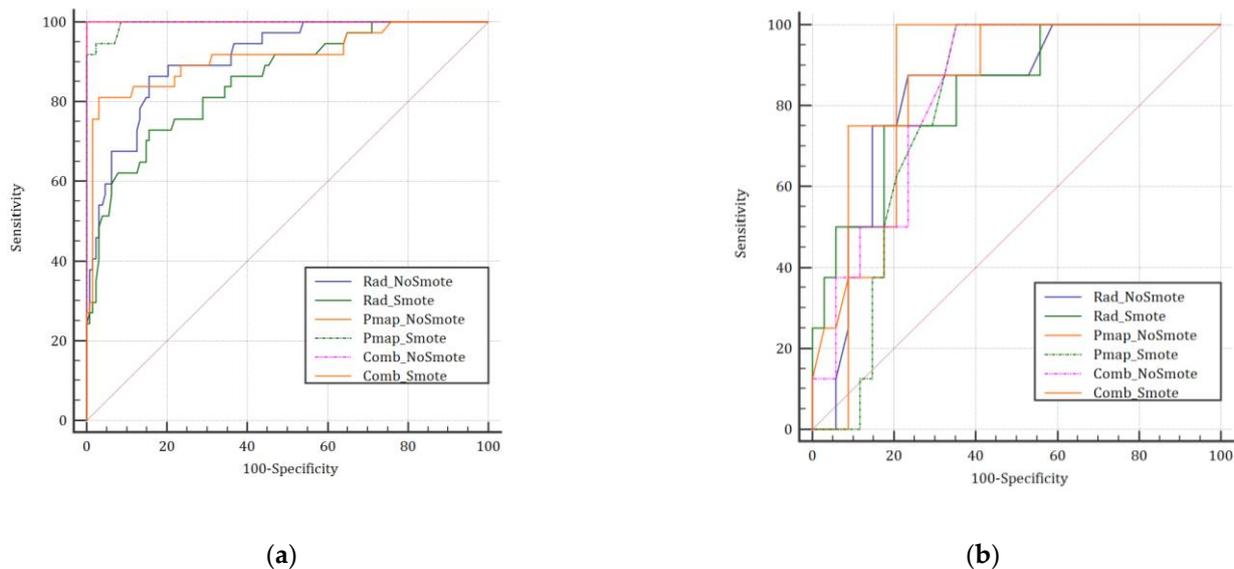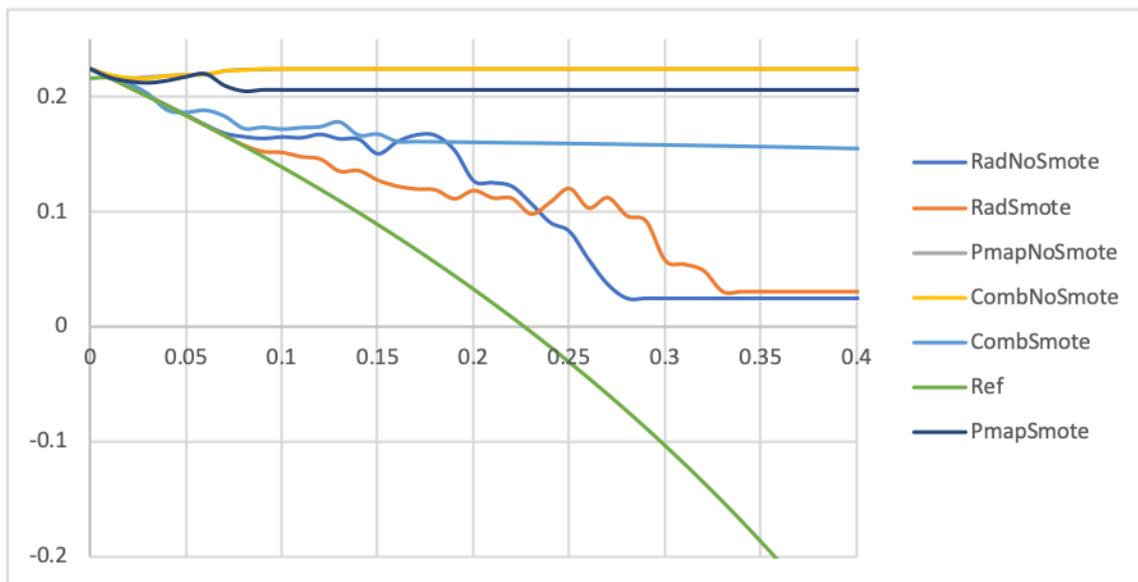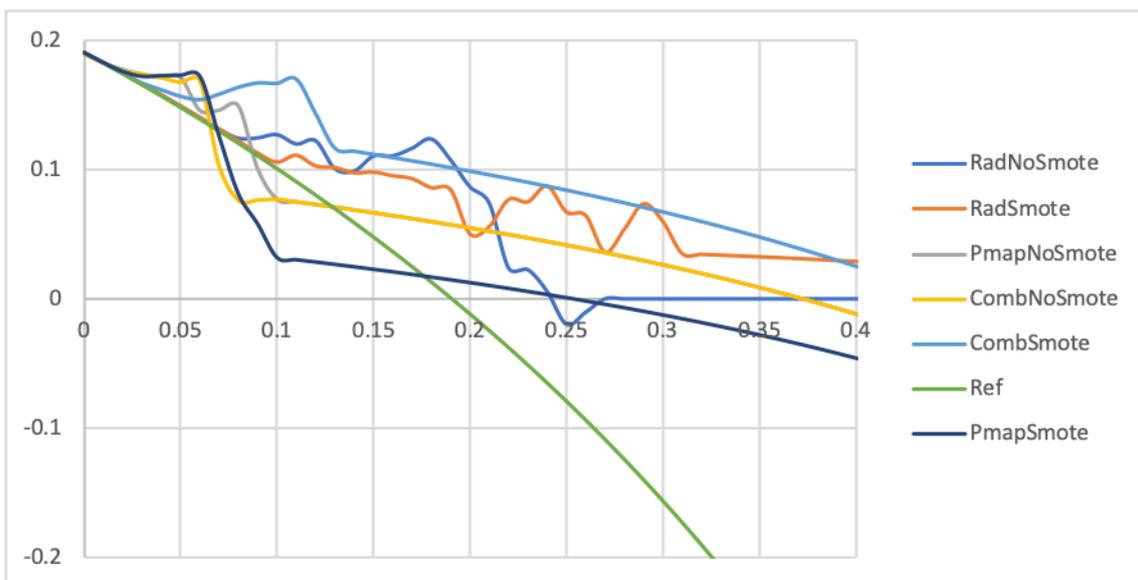

(**a**)   (**b**)

Figure 2. Comparison between each model in the training (**a**) and testing (**b**) sets based on the ROC curve for the prediction of APT ≥ grade 2. Abbreviation: Rad_NoSmote: Radiomics model without Smote, Rad_Smote: Radiomics model with Smote, Pmap_NoSmote: Pmap model without Smote, Pmap_Smote: Pmap model with Smote, Comb_NoSmote: Combined model without Smote, Comb_Smote: Combined model with Smote.

Though both the six-prediction models achieved interesting results in the testing set, the $Comb_{Smote}$ model achieved better results than the $Comb_{NoSmote}$, Rad, and Pmap models for the overall range of predicted probabilities (Figure 3). Similarly, the F1 score for the $Comb_{Smote}$ model surpassed all other models in the testing cohort (0.71 vs. 0.56–0.63). Precision-recall and calibration curves favored Smote-based models, except for the Pmap models (Supplementary Figures S2 and S3, respectively).

(**a**)

(**b**)

Figure 3. Comparison between each model in the training (**a**) and testing (**b**) sets based on the decision curve analysis for the prediction of APT $\geq$ grade 2. Abbreviation: APT: Acute pulmonary toxicity, Rad_NoSmote: Radiomics model without Smote, Rad_Smote: Radiomics model with Smote, Pmap_NoSmote: Pmap model without Smote, Pmap_Smote: Pmap model with Smote, Comb_NoSmote: Combined model without Smote, Comb_Smote: Combined model with Smote. Legend: x-axis: Probability threshold, y-axis: Net clinical benefit.

In both cohorts, according to the DCA, the radiomics model achieved the lowest clinical benefit for predicted probabilities >20%. On the other hand, for predicted probabilities <20%, net clinical benefit seemed to overlap in the testing cohort.

Regarding the prediction of APT $\geq$ grade 3 in the training cohort, three prediction models ranked first ($Pmap_{NoSmote}$, $Pmap_{Smote}$, and $Comb_{NoSmote}$), with a Bacc of 0.92 and newly defined respective cut-offs (Supplementary Table S4, Supplementary Figure S4a). On the testing cohort and with a 86% threshold, the $Pmaps_{Smote}$ model ranked first, with a Bacc of 0.94, a NPV

of 100%, but a low PPV of 28.6% (Supplementary Table S5, Supplementary Figure S4b). Apart from the Rad$_{NoSmote}$, all remaining models performed similarly with Baccs ranging from 0.85 to 0.93 and a NPV of 100.0% for all. Based on the DCA, the Pmap$_{Smote}$ and Comb$_{NoSmote}$ models harbored a favorable profile in the training cohort (Supplementary Figure S5a), whereas the Pmap$_{Smote}$ stood out in the testing cohort (Supplementary Figure S5b). To be noted, clinical benefits remained relatively small for all prediction models.

4. Discussion

Stage III lung cancer is a unique setting, where a tailored approach regarding response to treatment and risk of adverse events is mandatory. Indeed, though immunotherapy (IT) has changed the landscape of patients with locally advanced lung cancer [3,17], access to these adjuvant treatments must be increased. Furthermore, several phase II and phase III trials are currently ongoing to evaluate the benefit of IT with Ch-RT. In a recent Phase II trial [18] evaluating two different Ch-IT combinations with RT, 67 out of 112 patients treated with carboplatin—paclitaxel—pembrolizumab (patients with non-squamous or squamous lung cancer) discontinued treatment, 62.2% because of adverse events. Though objective response rates were interesting (approximately 70% in each arm), radiation pneumonitis of any grade was observed in 17.9% of the patients. Technical features regarding dosimetry planning and dose–volume histograms for both the PTV and OARs were insufficiently detailed in prospective trials evaluating treatments in this setting. Several studies assessing the compliance of the delivered planning to the protocols proved that survival was positively impacted by both volume definition and dose constraints [19,20]. One might think the initial dosimetry planning could also impact patients' tolerance to treatment. Given the number of trials evaluating Ch-IT with RT combinations [18,21–26], we developed two different approaches.

Our first approach considered the impact of spatial dosimetry and heterogeneity of the dose received by both lungs [5]. In the testing cohort, the Rad$_{NoSmote}$ model retrained on 165 patients achieved a lower Bacc than previously reported with a Bacc of 82.0% in the testing set. Based on only six radiomics features, this model is easily applicable but can only be considered as a method for evaluating dosimetry plans.

With a different approach based on voxel-based analysis, a cluster of voxels sensitive to radiotherapy was localized in the posterior right lung. Combining 11 features, the Pmap$_{NoSmote}$ model achieved similar results on the testing set (Bacc of 82.0%) compared with the radiomics model. Interestingly, the DCA favored the Pmap model for an APT$_2$ predicted probability superior to 20%. On the contrary, for probabilities inferior to 20%, performances of the two models overlapped. These two approaches are relatively innovative but have never been combined.

In the combined approach, we considered 12 features, among which the DMean$_{Pmap}$ and a radiomics feature (LungH_Variance) were associated. The Comb$_{NoSmote}$ model was associated with an AUC of 0.82 and Bacc of 69.5%, thus a lower Bacc mainly due to lower sensitivity. Addition of the SMOTE approach enhanced the performance of three newly developed prediction models. With a NPV of 100%, the rate of false positives decreased from 53.3% for the Rad$_{NoSmote}$ and Pmap$_{NoSmote}$ models to 46.7%. This model consisting of a two-step approach probably better apprehended the complexity of dose delivery to a functionally heterogenous organ such as the lung. The first step, consisting of an evaluation of the DMean$_{Pmap}$ and other clinical and dosimetric features, offers the possibility of adaptative planning with a goal of decreasing the DMean$_{Pmap}$. Once the dosimetry planning is completed, radiomics features are extracted and combined with previous features to produce a risk probability. A few advantages were found, including better results for APT$_2$ and APT$_3$ prediction. On the other hand, this approach remains more complex and less easily implemented for a clinical use. External validation of our models is currently under investigation. Despite validation in a prospectively recruited cohort, multicentric validation is needed. Several factors, such as treatment delivery heterogeneity, toxicity evaluation, and racial and ethnic differences, as well as psychosocial

factors, which could not be assessed in our cohorts, should be added to fully evaluate model performances.

Apart from the Rad_{Smote} and $Comb_{Smote}$, all models were prone to overfitting with significant performance loss between training and testing cohorts. Rad_{Smote} and $Comb_{Smote}$ achieved harbored correction for imbalanced data that seemed to reduce this limitation and should be considered in future prediction models when dealing with skewed endpoints. The $Comb_{Smote}$ model also stood out from the other models with the highest F1 score.

Having more efficient APT prediction tools could greatly benefit patients undergoing RT for lung cancer. Patients at high predicted risk of APT should be considered for dosimetry optimization until an acceptable risk is reached. When unfeasible, careful monitoring under treatment or neoadjuvant chemotherapy could be proposed. Patients at low risk of APT could benefit from dose escalation or treatment combination, with the addition of IT during CT-RT followed by adjuvant IT, as studied in several phase III trials [18,21–26].

Main alternatives to such machine-learning approaches are based on nuclear imaging and require costly exams, such as ventilation (V)/perfusion (P), SPECT-CT, or ^{68}Ga PET-CT [6–8]. These approaches have the advantage of offering the possibility of adaptative planning, based on the physiological structure of the lungs. However, several technical issues arise, such as 4D inter-modality registration and variability of the definitions of functional and non-functional lungs. Furthermore, the possibility of effective optimization of dosimetry planning must be demonstrated. Such approaches could further enhance prediction modeling and treatment planning.

Though the Pmap and PmapRad volumes appeared complementary, both regions were mainly encompassed in the right lung, confirming the specific role of these sub-regions regarding the risk of APT. Despite an overlap between these regions and functional regions, as defined through nuclear imaging [6], the precise physiopathology underlying the role of these regions needs further research. Specific vascularization of the posterior right lung could be a primary explanation. Sensitivity of these regions to low-dose volumes ($V5_{LungH}$ and $V10_{LungH}$) and their impacts on respiratory functions, especially on the diffusion capacity of the lungs for carbon monoxide (DLCO), could partly explain our results. A phase 2 study (NCT04942275) being conducted at our institution focuses on patients treated with stereotactic radiotherapy (SBRT), aimed at better understanding regional radiation sensitivity [27]. Despite it being on a different population, substantial information will be extracted from this study and may explain the specific role of the posterior right lung.

Despite its few limitations, this study brings new insight for prediction modelling in patients treated with radiotherapy and marks the final step before clinical validation in a randomized setting. With two competitive models, comparison and combination appear necessary to choose the best model for validation. This choice is based on prediction efficiency, but also generalizability, ease of use, and integration in the usual workflow, with the Pmap model being the most balanced model. This manuscript stands out from previous studies due to several factors. Due to the relatively moderate cohort and high number of features, a correction of imbalanced data was performed, enhancing the overall results for several models. As hypothetized, all non-SMOTE models suffered from overfitting. This was the main benefit of the SMOTE approach, with enhanced performance for the $Comb_{Smote}$ model regarding the prediction of grade > 2 APT. Finally, this work brought a focus on APT_3 prediction with, again, the combination model reaching the highest performances.

5. Conclusions

The accuracy of the APT2 prediction models is increased by combining radiomics, DMeanPmap, and common clinical and dosimetric features. The combined model appears to be more effective for the prediction of APT $\geq$ grades 2 and 3 than other prediction models, such as the Pmap-model in patients treated with VMAT for locally advanced lung cancer. These two models are useful tools for clinicians to reduce the risk of APT in selected patients and are ready for prospective and external validation in a multicentric setting. Following this validation, a randomized study will be initiated comparing outcomes in

patients treated with dosimetric optimization based on the previously presented Pmap model and patients treated without dosimetric optimization. Such a randomized trial will need a simple, efficient, and generalizable prediction model that would be either embedded in the treatment planning system or act as a standalone tool. Furthermore, the use of this model may provide access to novel therapeutic alternatives, such as dose escalation or creative therapy combinations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpm12111926/s1, Supplementary Table S1: List of all studied dosimetric features, Supplementary Table S2: Main patient characteristics in the training and testing sets, Supplementary Table S3: Correlation of each feature with the APT_2 risk (training cohort), Supplementary Table S4: Analysis of the each model's discrimination between patients with or without APT $\geq$ grade 3 in the training set, Supplementary Table S5: Analysis of the each model's discrimination between patients with or without APT $\geq$ grade 3 in the testing set, Supplementary Figure S1: Overlap between the PmapRad and Pmap maps, Supplementary Figure S2: Comparison between each model in the testing sets based on the precision-recall curve for the prediction of APT $\geq$ grade 2, Supplementary Figure S3: Comparison between each model in the testing sets based on the calibration curve for the prediction of APT $\geq$ grade 2, Supplementary Figure S4: Comparison between each model in the training (a) and testing (b) sets, based on the ROC curve for the prediction of APT $\geq$ grade 3, Supplementary Figure S5: Comparison between each model in the training (a) and testing (b) sets, based on the decision curve analysis for the prediction of APT $\geq$ grade 3.

References

1. Amin, M.B.; Greene, F.L.; Edge, S.B.; Compton, C.C.; Gershenwald, J.E.; Brookland, R.K.; Meyer, L.; Gress, D.M.; Byrd, D.R.; Winchester, D.P. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J. Clin.* **2017**, *67*, 93–99. [CrossRef] [PubMed]
2. Ettinger, D.S.; Wood, D.E.; Aisner, D.L.; Akerley, W.; Bauman, J.R.; Bharat, A.; Bruno, D.S.; Chang, J.Y.; Chirieac, L.R.; D'Amico, T.A.; et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 2.2021. *J. Natl. Compr. Cancer Netw.* **2021**, *19*, 254–266. [CrossRef] [PubMed]
3. Antonia, S.J.; Villegas, A.; Daniel, D.; Vicente, D.; Murakami, S.; Hui, R.; Yokoi, T.; Chiappori, A.; Lee, K.H.; De Wit, M.; et al. Durvalumab after Chemoradiotherapy in Stage III Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2017**, *377*, 1919–1929. [CrossRef]
4. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; 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] [PubMed]
5. Bourbonne, V.; Da-Ano, R.; Jaouen, V.; Lucia, F.; Dissaux, G.; Bert, J.; Pradier, O.; Visvikis, D.; Hatt, M.; Schick, U. Radiomics analysis of 3D dose distributions to predict toxicity of radiotherapy for lung cancer. *Radiother. Oncol.* **2021**, *155*, 144–150. [CrossRef]
6. Bucknell, N.; Hardcastle, N.; Jackson, P.; Hofman, M.; Callahan, J.; Eu, P.; Iravani, A.; Lawrence, R.; Martin, O.; Bressel, M.; et al. Single-arm prospective interventional study assessing feasibility of using gallium-68 ventilation and perfusion PET/CT to avoid functional lung in patients with stage III non-small cell lung cancer. *BMJ Open* **2020**, *10*, e042465. [CrossRef]
7. Le Roux, P.Y.; Hicks, R.J.; Siva, S.; Hofman, M.S. PET/CT Lung Ventilation and Perfusion Scanning Using Galligas and Gallium-68-MAA. *Semin. Nucl. Med.* **2019**, *49*, 71–81. [CrossRef]

8. Siva, S.; Thomas, R.; Callahan, J.; Hardcastle, N.; Pham, D.; Kron, T.; Hicks, R.; MacManus, M.; Ball, D.L.; Hofman, M.S. High-resolution pulmonary ventilation and perfusion PET/CT allows for functionally adapted intensity modulated radiotherapy in lung cancer. *Radiother. Oncol.* **2015**, *115*, 157–162. [CrossRef]

9. Westergaard, S.A.; Castillo, R.; Rupji, M.; Bell, J.; Castro, R.; Harms, J.; Kayode, O.; Nguyen, T.; Wolf, J.; Higgins, K.; et al. Sparing Functional Lung Using Optimized Volumetric Modulated Arc Therapy in Thoracic Radiation for Locally Advanced Non-Small Cell Lung Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2021**, *111*, e456–e457. [CrossRef]

10. Palma, G.; Monti, S.; Cella, L. Voxel-based analysis in radiation oncology: A methodological cookbook. *Phys. Med.* **2020**, *69*, 192–204. [CrossRef]

11. Palma, G.; Monti, S.; Xu, T.; Scifoni, E.; Yang, P.; Hahn, S.M.; Durante, M.; Mohan, R.; Liao, Z.; Cella, L. Spatial Dose Patterns Associated with Radiation Pneumonitis in a Randomized Trial Comparing Intensity-Modulated Photon Therapy with Passive Scattering Proton Therapy for Locally Advanced Non-Small Cell Lung Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2019**, *104*, 1124–1132. [CrossRef]

12. Bourbonne, V.; Lucia, F.; Jaouen, V.; Bert, J.; Rehn, M.; Pradier, O.; Visvikis, D.; Schick, U. Development and prospective validation of a spatial dose pattern based model predicting acute pulmonary toxicity in patients treated with volumetric arc-therapy for locally advanced lung cancer. *Radiother. Oncol.* **2021**, *164*, 43–49. [CrossRef] [PubMed]

13. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [CrossRef] [PubMed]

14. Lema, G.; Nogueira, F.; Aridas, C.K. Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.

15. 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.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef]

16. Bourbonne, V.; Delafoy, A.; Lucia, F.; Quéré, G.; Pradier, O.; Schick, U. Toxicity after volumetric modulated arc therapy for lung cancer: A monocentric retrospective study. *Transl. Lung Cancer Res.* **2021**, *10*, 156–166. [CrossRef]

17. Antonia, S.J.; Villegas, A.; Daniel, D.; Vicente, D.; Murakami, S.; Hui, R.; Kurata, T.; Chiappori, A.; Lee, K.H.; De Wit, M.; et al. Overall Survival with Durvalumab after Chemoradiotherapy in Stage III NSCLC. *N. Engl. J. Med.* **2018**, *379*, 2342–2350. [CrossRef]

18. Jabbour, S.K.; Lee, K.H.; Frost, N.; Breder, V.; Kowalski, D.; Alawin, I.; Levchenko, E.; Reguart, N.; Martinez-Marti, A.; Houghton, B.; et al. Pembrolizumab Plus Platinum Chemotherapy and Radiotherapy for Unresectable, Locally Advanced, Stage 3 NSCLC: KEYNOTE-799. *Int. J. Radiat. Oncol. Biol. Phys.* **2021**, *111*, S9–S10. [CrossRef]

19. Weber, D.C.; Hurkmans, C.W.; Melidis, C.; Budach, W.; Langendijk, J.H.; Peters, L.J.; Gregoire, V.; Maingon, P.; Combescure, C. Outcome impact and cost-effectiveness of quality assurance for radiotherapy planned for the EORTC 22071–24071 prospective study for head and neck cancer. *Radiother. Oncol.* **2014**, *111*, 393–399. [CrossRef]

20. Ohri, N.; Shen, X.; Dicker, A.; Doyle, L.A.; Harrison, A.S.; Showalter, T. Radiotherapy Protocol Deviations and Clinical Outcomes: A Meta-analysis of Cooperative Group Clinical Trials. *J. Natl. Cancer Inst.* **2013**, *105*, 387–393. [CrossRef]

21. Peters, S.; Felip, E.; Dafni, U.; Belka, C.; Guckenberger, M.; Irigoyen, A.; Nadal, E.; Becker, A.; Vees, H.; Pless, M.; et al. Safety evaluation of nivolumab added concurrently to radiotherapy in a standard first line chemo-radiotherapy regimen in stage III non-small cell lung cancer—The ETOP NICOLAS trial. *Lung Cancer* **2019**, *133*, 83–87. [CrossRef] [PubMed]

22. Lin, S.H.; Lin, Y.; Yao, L.; Kalhor, N.; Carter, B.W.; Altan, M.; Blumenschein, G.; Byers, L.A.; Fossela, F.; Gibbons, D.L.; et al. Phase II Trial of Concurrent Atezolizumab with Chemoradiation for Unresectable NSCLC. *J. Thorac. Oncol.* **2020**, *15*, 248–257. [CrossRef] [PubMed]

23. Jabbour, S.K.; Cho, B.C.; Bria, E.; Kato, T.; Bhosle, J.; Gainor, J.; Reguart, N.; Wang, L.; Morgensztern, D.; Gurary, E.; et al. KEYLYNK-012: A Phase 3 Study of Pembrolizumab with Concurrent Chemoradiation Therapy (CCRT) Followed by Pembrolizumab with or without Olaparib vs. CCRT Followed by Durvalumab in Unresectable, Locally Advanced, Stage III Non-Small-Cell Lung Cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **2021**, *111*, e440–e441. [CrossRef]

24. Bradley, J.D.; Nishio, M.; Okamoto, I.; Newton, M.D.; Trani, L.; Shire, N.J.; Gu, Y.; Dennis, P.A.; Lee, K.H. PACIFIC-2: Phase 3 study of concurrent durvalumab and platinum-based chemoradiotherapy in patients with unresectable, stage III NSCLC. *J. Clin. Oncol.* **2019**, *37*, TPS8573. [CrossRef]

25. Leal, T.; Wang, Y.; Dowlati, A.; Lewis, D.A.; Chen, Y.; Mohindra, A.R.; Razaq, M.; Ahuja, H.G.; Liu, J.; King, D.M.; et al. Randomized phase II clinical trial of cisplatin/carboplatin and etoposide (CE) alone or in combination with nivolumab as frontline therapy for extensive-stage small cell lung cancer (ES-SCLC): ECOG-ACRIN EA5161. *J. Clin. Oncol.* **2020**, *38*, 9000. [CrossRef]

26. Varlotto, J.M.; Sun, Z.; Ky, B.; Upshaw, J.; Katz, S.I.; Fitzgerald, T.J.; Wakelee, H.; Diehn, M.; Mankoff, D.A.; Lovely, C.; et al. A Review of Immunotherapy for Stage III and Metastatic Non-Small Cell Lung Cancer and the Rationale for the ECOG-ACRIN EA5181 Study. *Oncologist* **2021**, *26*, 523–532. [CrossRef]

27. Lucia, F.; Rehn, M.; Blanc-Béguin, F.; Le Roux, P.-Y. Radiation Therapy Planning of Thoracic Tumors: A Review of Challenges Associated with Lung Toxicities and Potential Perspectives of Gallium-68 Lung PET/CT Imaging. *Front. Med.* **2021**, *8*, 723748. [CrossRef]