

## Article

# Predictive Value of $^{18}\text{F}$ -FDG PET/CT Using Machine Learning for Pathological Response to Neoadjuvant Concurrent Chemoradiotherapy in Patients with Stage III Non-Small Cell Lung Cancer

Jang Yoo <sup>1</sup>, Jaeho Lee <sup>2</sup>, Miju Cheon <sup>1</sup>, Sang-Keun Woo <sup>3</sup>, Myung-Ju Ahn <sup>4</sup>, Hong Ryull Pyo <sup>5</sup>, Yong Soo Choi <sup>6</sup>, Jung Ho Han <sup>7</sup> and Joon Young Choi <sup>8,\*</sup>

- <sup>1</sup> Department of Nuclear Medicine, Veterans Health Service Medical Center, Seoul 05368, Korea; jang8214.yoo@gmail.com (J.Y.); diva1813@naver.com (M.C.)
  - <sup>2</sup> Department of Preventive Medicine, Seoul National University College of Medicine, Seoul 03080, Korea; hoyajh21@gmail.com
  - <sup>3</sup> Department of Nuclear Medicine, Korea Cancer Center Hospital, Korea Institute of Radiological and Medical Sciences (KIRAMS), Seoul 01812, Korea; skwoo@kirams.re.kr
  - <sup>4</sup> Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; silk.ahn@samsung.com
  - <sup>5</sup> Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; hr.pyo@samsung.com
  - <sup>6</sup> Department of Thoracic and Cardiovascular Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; ysooyah.choi@samsung.com
  - <sup>7</sup> Department of Pathology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea; joungho.han@samsung.com
  - <sup>8</sup> Department of Nuclear Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea
- \* Correspondence: jynm.choi@samsung.com; Tel.: +82-2-3410-2648; Fax: +82-2-3410-2639



**Citation:** Yoo, J.; Lee, J.; Cheon, M.; Woo, S.-K.; Ahn, M.-J.; Pyo, H.R.; Choi, Y.S.; Han, J.H.; Choi, J.Y. Predictive Value of  $^{18}\text{F}$ -FDG PET/CT Using Machine Learning for Pathological Response to Neoadjuvant Concurrent Chemoradiotherapy in Patients with Stage III Non-Small Cell Lung Cancer. *Cancers* **2022**, *14*, 1987. <https://doi.org/10.3390/cancers14081987>

Academic Editors: Hamid Khayyam, Ali Madani, Rahele Kafieh and Ali Hekmatnia

Received: 17 February 2022

Accepted: 13 April 2022

Published: 14 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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:** The pathological complete response (pCR) after neoadjuvant chemoradiotherapy (CCRT) is an independent prognostic factor for progression-free and overall survival in non-small cell lung cancer (NSCLC).  $^{18}\text{F}$ -FDG PET/CT has been performed for initial staging work-up, treatment response, and follow-up in patients with NSCLC. Machine learning (ML) as an empirical data science has become relevant to nuclear medicine. We investigated the predictive performance of  $^{18}\text{F}$ -FDG PET/CT using an ML model to assess the treatment response to neoadjuvant CCRT in patients with stage III NSCLC, and compared the performance of the ML model predictions to predictions from conventional PET parameters and from physicians. The predictions from the ML model using radiomic features of  $^{18}\text{F}$ -FDG PET/CT provided better accuracy than predictions from conventional PET parameters and from physicians for the neoadjuvant CCRT response of stage III non-small cell lung cancer.

**Abstract:** We investigated predictions from  $^{18}\text{F}$ -FDG PET/CT using machine learning (ML) to assess the neoadjuvant CCRT response of patients with stage III non-small cell lung cancer (NSCLC) and compared them with predictions from conventional PET parameters and from physicians. A retrospective study was conducted of 430 patients. They underwent  $^{18}\text{F}$ -FDG PET/CT before initial treatment and after neoadjuvant CCRT followed by curative surgery. We analyzed texture features from segmented tumors and reviewed the pathologic response. The ML model employed a random forest and was used to classify the binary outcome of the pathological complete response (pCR). The predictive accuracy of the ML model for the pCR was 93.4%. The accuracy of predicting pCR using the conventional PET parameters was up to 70.9%, and the accuracy of the physicians' assessment was 80.5%. The accuracy of the prediction from the ML model was significantly higher than those derived from conventional PET parameters and provided by physicians ( $p < 0.05$ ). The ML model is useful for predicting pCR after neoadjuvant CCRT, which showed a higher predictive accuracy than those achieved from conventional PET parameters and from physicians.

**Keywords:** non-small cell lung cancer; neoadjuvant concurrent chemoradiotherapy;  $^{18}\text{F}$ -FDG PET/CT; machine learning; random forest; pathologic complete response

## 1. Introduction

Lung cancer is the most common malignant tumor and remains the leading cause of cancer-related death worldwide in spite of major advances in prevention and multimodal treatment [1]. Non-small cell lung cancer (NSCLC) accounts for more than 85% of all lung cancers and about 30% of NSCLC present with locally advanced disease in stage III [2]. Patients with stage III NSCLC are usually considered as inoperable. Neoadjuvant concurrent chemoradiotherapy (CCRT) followed by surgery has been established as being able to improve the overall outcome by reducing the rate of local failures and distant metastasis [3,4].

In patients receiving neoadjuvant CCRT for stage III NSCLC, surgical resection allows for the identification of the histopathologic tumor response to determine the prognosis and to evaluate postoperative therapeutic options. According to previous studies, the pathologic complete response (pCR) after neoadjuvant CCRT is an independent prognostic factor for progression-free and overall survival in NSCLC [5,6]. Although several papers have reported a wide range of pCR values of 16–27%, it is clear that the pCR is highly correlated with patient survival [7–10].

$^{18}\text{F}$ -fluorodeoxyglucose positron emission tomography/computed tomography ( $^{18}\text{F}$ -FDG PET/CT) has been performed for initial staging work-up, treatment response, and follow-up in patients with NSCLC. It has also been viewed as appropriate for the precise investigation of treatment response after CCRT [11,12]. Previous studies have focused on the comparison of quantitative PET parameters such as the standard uptake value (SUV) after neoadjuvant treatment and histopathologic findings after surgery [13,14]. Moreover, the application of the PET response criteria in solid tumors (PERCIST 1.0) as an evaluation for  $^{18}\text{F}$ -FDG PET/CT has been performed to enhance the limitation of anatomic tumor response metrics [15,16]. The role of  $^{18}\text{F}$ -FDG PET/CT still needs to be explored because possible misinterpretations due to radiation-induced inflammation such as pneumonitis can cause problems in  $^{18}\text{F}$ -FDG PET/CT images [17,18].

Machine learning (ML) as an empirical data science, which can learn patterns or characteristics from one set of given data and use them to evaluate new data, has become relevant to nuclear medicine. Our previous study demonstrated that ML is well suited to performing analyses of high dimensionality radiomic feature extraction from  $^{18}\text{F}$ -FDG PET/CT, and ML analysis provided better diagnostic performance than physicians for evaluating metastatic mediastinal lymph nodes in NSCLC [19]. Although assessing the radiomic features of a tumor in clinical practice has some challenges because of the time, effort, and skill involved, we have shown that ML can improve the diagnostic accuracy and its availability in NSCLC. However, there is still no study that has evaluated the predictive performance of ML for the neoadjuvant CCRT response using the radiomic features of  $^{18}\text{F}$ -FDG PET/CT.

Therefore, we investigated the predictive performance of  $^{18}\text{F}$ -FDG PET/CT using an ML model to assess the treatment response to neoadjuvant CCRT in patients with stage III NSCLC, and compared the performance of the ML model predictions to predictions from conventional PET parameters and from physicians.

## 2. Materials and Methods

### 2.1. Subjects

We retrospectively reviewed the medical records of all patients newly diagnosed with stage III NSCLC through imaging studies such as chest X-ray, enhanced chest CT, and  $^{18}\text{F}$ -FDG PET/CT, as well as pathologic studies including endobronchial ultrasound-guided transbronchial needle aspiration, mediastinoscopic biopsy, or thoracotomy, between

November 2008 and October 2020. To be included in the study population, patients needed to complete a planned neoadjuvant CCRT and undergo curative-intent surgical treatment for stage III NSCLC according to the 7th edition of the TNM classification [20], and undergo a second  $^{18}\text{F}$ -FDG PET/CT within approximately 3 weeks following the completion of neoadjuvant CCRT for restaging work-up. Patients in poor cardiopulmonary condition that precluded surgery or who had previously been treated because of another malignant disease were excluded from the study population. Patients who received neoadjuvant chemotherapy or radiotherapy alone were also excluded.

This study was approved by the institutional review board of our institution (IRB No. 2020-09-185), and the requirement for informed patient consent was waived due to its retrospective design.

## 2.2. Neoadjuvant CCRT and Histopathologic Evaluation

The neoadjuvant CCRT consisted of chemotherapy and concurrent thoracic radiotherapy. Thoracic radiotherapy was delivered to patients with a total dose of 45 Gy with 1.8 Gy/fraction over 5 weeks from November 2008 to October 2009 or 44 Gy with 2.0 Gy/fraction over 4.5 weeks using 10-MV X-rays from October 2009 and thereafter. The radiotherapy target volume included the known gross and clinical disease plus adequate peripheral margins. The chemotherapy regimens mostly consisted of intravenous administration of paclitaxel (50 mg/m<sup>2</sup> per week) or docetaxel (20 mg/m<sup>2</sup> per week) plus either cisplatin (25 mg/m<sup>2</sup> per week) or carboplatin (AUC, 1.5/week) for 5 weeks. The first dose of chemotherapy was delivered on the first day of thoracic radiotherapy [3,4,21].

Surgical procedures were planned for 4–6 weeks following the completion of neoadjuvant CCRT and comprised resection of the affected lung plus mediastinal lymph nodes dissection, depending on the clinical stage. Pulmonary resection included lobectomy, bilobectomy, pneumonectomy, or lobectomy with en bloc wedge resection according to the extent of the primary tumor. After surgical resection, the specimens were examined by pathologists for residual tumors based on hematoxylin and eosin-stained slides. They reported the percentage of residual tumor, which was determined by comparing the estimated cross-sectional area of the viable tumor foci with the estimated cross-sectional areas of necrosis, fibrosis, and inflammation on each slide. The absolute viable tumor extent was also assessed based on their calculation, and pathologic complete response (pCR) was defined as no residual viable tumor remaining in the post-therapy pathology specimen [22,23].

## 2.3. $^{18}\text{F}$ -FDG PET/CT Analysis

All patients fasted for at least 6 h before  $^{18}\text{F}$ -FDG PET/CT was performed to keep their blood glucose level below 200 mg/dL. Torso PET and unenhanced CT images were acquired using a dedicated PET/CT scanner (Discovery STe, GE Healthcare, Waukesha, WI, USA) approximately 60 min after intravenous injection of 5.5 MBq/kg of  $^{18}\text{F}$ -FDG. CT images were obtained using a 16-slice helical CT with the following settings: 140 keV, 30–170 mAs with Auto A mode, and a slice section of 3.75 mm. PET images were acquired from head to thigh and attenuation-corrected PET images (voxel size, 3.9 × 3.9 × 3.3 mm<sup>3</sup>) were reconstructed using a 3D ordered-subset expectation-maximization algorithm (20 subsets, 2 iterations).

For quantitative analysis, the volume of interest (VOI) from the primary tumor was delineated using the gradient-based segmentation method (PET Edge) in MIM version 6.4 (MIM Software Inc., Cleveland, OH, USA) [19]. These VOIs were saved as a DICOM-RT structure that was imported into the Chang-Gung Image Texture Analysis toolbox (CGITA, <http://code.google.com/p/cgita>, accessed on 1 March 2020) facilitated by MATLAB software (version 2014b; MathWorks, Inc., Natick, MA, USA) to extract the radiomic features from the PET images (Supplemental Table S1) as well as conventional PET parameters, including the maximum SUV (SUV<sub>max</sub>), mean SUV (SUV<sub>mean</sub>), metabolic tumor volume (MTV), and total lesion glycolysis (TLG). We also calculated the differences of these con-

ventional parameters between PET1 and PET2 by subtracting PET2 parameters from those of PET1 and dividing by those of PET1.

Two nuclear medicine physicians (J.Y.C. and B.T.K) with more than 15 years of experience in PET/CT interpretation assessed the neoadjuvant treatment response according to PERCIST 1.0 [16] by means of a baseline  $^{18}\text{F}$ -FDG PET/CT (PET1) and second PET/CT (PET2) undertaken before surgery. They categorized all patients into four response criteria: complete metabolic response (CMR), partial metabolic response (PMR), stable metabolic disease (SMD), and progressive metabolic disease (PMD). After that, the accuracy of the predicted CMR results were compared to histopathologic pCR.

#### 2.4. Machine Learning (ML) Model

The ML model was developed as a binary classification. First, data were partitioned into a training dataset (70%) for model building and an independent testing dataset (30%) for internal validation. We developed an ML tree-based boosting model for pCR prediction using a random forest (RF) algorithm, which consisted of a multitude of decision trees and used an ensemble method to decide the outcome. Our model was trained with the bagging method to predict the pCR. Different numbers of trees were used to classify the binary decision of the result to achieve the best performance score. The Gini impurity was measured to the quality of a split. The maximum depth of the tree was 5, and the square root of the number of the features was considered for the max. number of features to look for the best split of the model. We applied a random grid search method to determine the optimal hyperparameter of the RF model [24–27]. A 10-fold cross-validation in the training dataset, a technique for reducing the bias that can occur as a result of using a single training set, was applied for method validation. All ML statistical analyses were performed using Python (version 3.8.3).

In classic oversampling techniques, the minority data are simply replicated from the minority data population. The ML model does not reflect on variation from the oversampling data. Therefore, we tried to use SMOTE (Synthetic Minority Oversampling Technique) to deal with this class problem. This technique helped with unbalanced data by creating new synthetic data to provide balance in the distribution. SMOTE starts by choosing random data from the minority class. Then it uses a K-Nearest Neighbor (KNN) algorithm to set new points of the data. Next, new synthetic data are created between the random data and new point, which is derived from KNN algorithm. This process is repeated until the minority class reaches the same size as the majority class. Therefore, we added 322 more participants from the existing raw data. A total of 752 participants were analyzed using this oversampling technique.

Several useful scaling techniques (Min–Max scaler, Normalization, Standardization) prevent overflow and underflow of the data. They help to compare dimensional data more efficiently through a scaling process. The process reduces the conditional number of covariance matrices from the independent variables. This reduction enhances the speed of conversion and stability of the model during the optimization process. We used a standard scaler, which removes the mean and helps to scale the value's unit variance. To adjust for the different scales of the features, standardization of the variables is necessary for the preprocessing steps.

For feature selection, top 10, 20, and 30 variables among 144 variables were selected according to the importance of the variables based on the mean decrease impurity (MDI). MDI or Gini importance was calculated as the decrease in node impurity weighted by the probability of reaching the node. The sum over the number of splits decided the variable importance of the model. The higher value of MDI meant the critical feature in the model.

#### 2.5. Statistical Analysis

The association between conventional PET parameters and pCR was determined by an independent *t*-test or the Mann–Whitney test according to the Kolmogorov–Smirnov test. Receiver operating characteristic (ROC) curve analysis was performed to assess

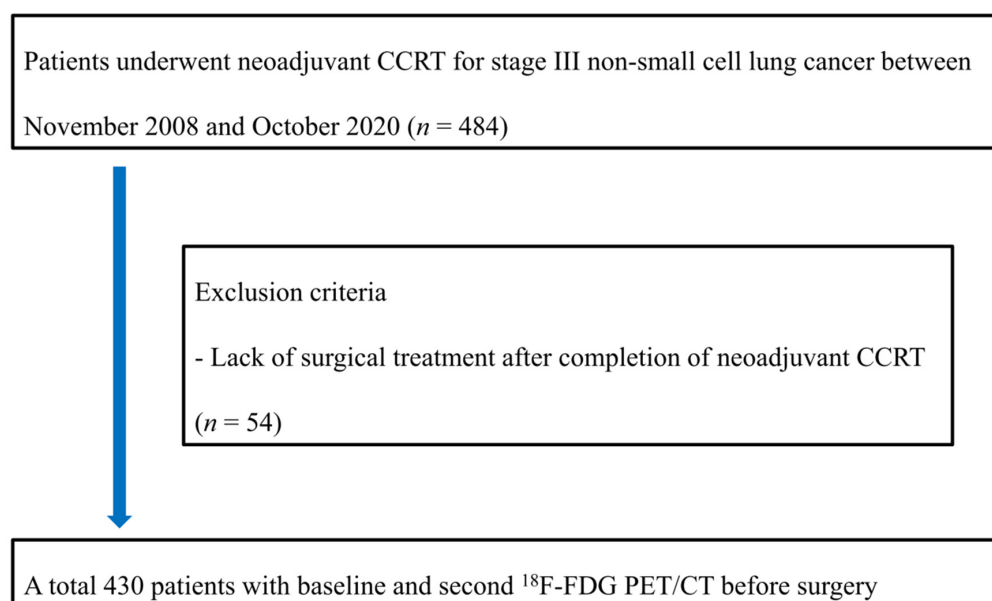
optimal cutoff values of continuous variables using the MedCalc software package (Ver. 9.5, MedCalc Software, Mariakerke, Belgium). The predictive performance of conventional PET parameters and physicians' diagnostic results were reported using sensitivity (Sen), specificity (Spe), positive predictive value (PPV), negative predictive value (NPV), and accuracy (ACC).

For predictive performance of the ML model, we measured the areas under curve (AUCs), ACC, F1 score, precision (also called PPV), and recall (also known as Sen). We compared the measured values with those of predictions from conventional PET parameters and from physicians by using a McNemar test or Fisher's exact test. A  $p$ -value of less than 0.05 was considered statistically significant.

### 3. Results

#### 3.1. Subject Characteristics

Among 484 consecutive patients, 430 patients were enrolled in this study. Fifty-four patients were excluded from the analysis due to a lack of surgical treatment after completion of neoadjuvant CCRT (Figure 1). The clinical characteristics of the 430 patients are summarized in Table 1. The patients were predominantly male (71.9%), and there was a high prevalence (67.2%) of adenocarcinoma among the patients. After neoadjuvant CCRT followed by surgery, the mean percentage of viable tumor in the pathologic specimen was 28.8% (range 0–95%). The pCR was observed in 54 patients (12.6%). According to PERCIST criteria, 16.7% of patients had CMR ( $n = 72$ ).



**Figure 1.** Flowchart of the inclusion and exclusion criteria for the patients.

#### 3.2. Predictive Performance of ML Model for pCR

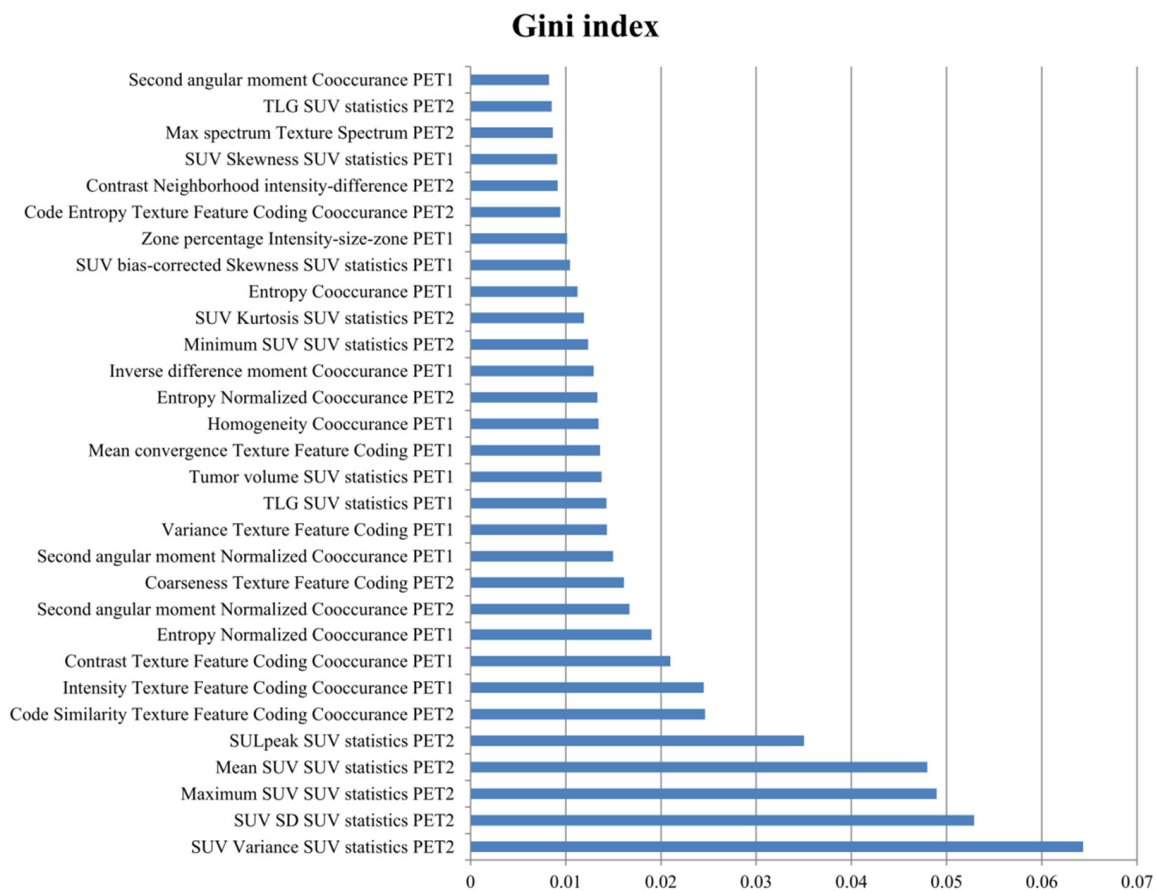
The radiomic feature importance was obtained using a Gini index representing the coefficient of the attributes on the prediction model, as listed in Figure 2. The overall prediction performance of the ML model was compared by calculating each of the PET1 and PET2 features separately, and all variables from both PET1 and PET2 (PET3) were analyzed (Table 2). The AUCs determined by the ML model were 0.934 in PET1, 0.975 in PET2, and 0.977 in PET3. For comparison ROC curve analysis (Figure 3), the AUCs of PET2 and PET3 were significantly higher than that of PET1 ( $p = 0.009$ ,  $p = 0.006$ , respectively). However, there was no significant difference between the AUCs of PET2 and PET3 ( $p = 0.805$ ). According to other indices, PET3 revealed a better predictive performance than those results with either PET1 or PET2 variables.



**Table 1.** Subjects’ characteristics.

Characteristics	No.	
Sex	Male	309 (71.9%)
	Female	121 (28.1%)
Age (years)	Mean (range)	61.8 (31.1–79.5)
Tumor pathology	Adenocarcinoma	289 (67.2%)
	Squamous cell carcinoma	125 (29.1%)
	Others	16 (3.7%)
Stage (AJCC 7th)	IIIa	415 (96.5%)
	IIIb	15 (3.5%)
Type of surgery	Lobectomy	339 (78.8%)
	Bilobectomy	32 (7.4%)
	Pneumonectomy	23 (5.4%)
	Lobectomy with en bloc wedge resection	36 (8.4%)
Viable tumor on pathologic specimen	Mean % (range)	28.8 (0–95.0)
Pathologic response	pCR	54 (12.6%)
	Non-pCR	376 (87.4%)
Response by PERCIST	CMR	72 (16.7%)
	PMR	281 (65.4%)
	SMD	72 (16.7%)
	PMD	5 (1.2%)

pCR, pathologic complete response; PERCIST, PET response criteria in solid tumors; CMR, complete metabolic response; PMR, partial metabolic response; SMD, stable metabolic disease; PMD, progressive metabolic disease.

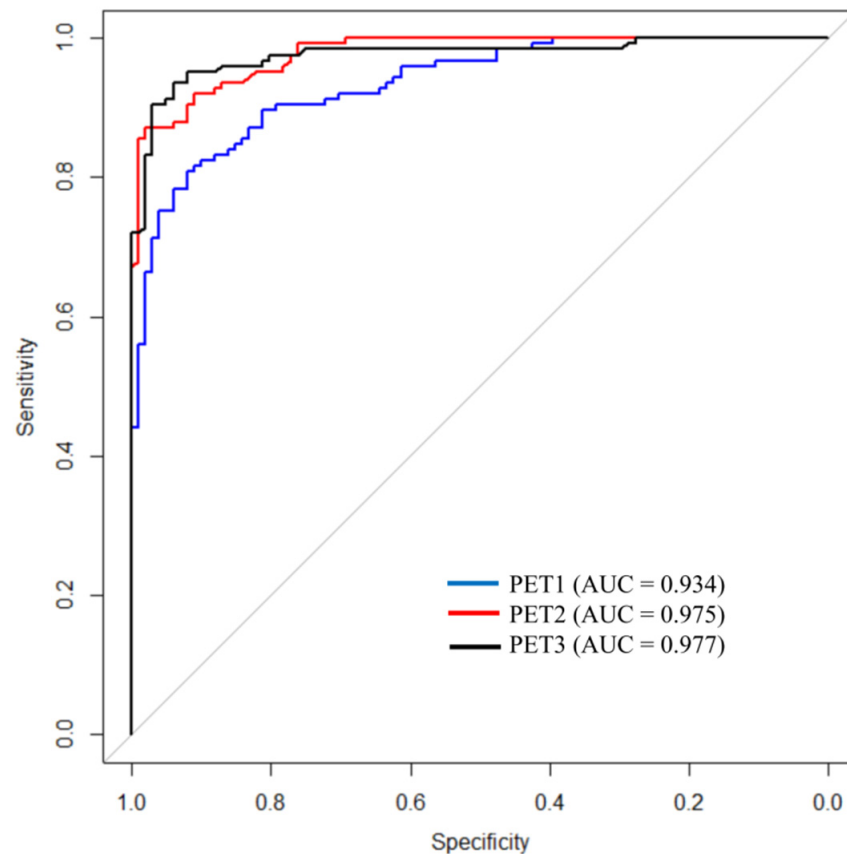


**Figure 2.** The top 30 important radiomic features from <sup>18</sup>F-FDG PET/CT for pCR prediction after neoadjuvant CCRT.

**Table 2.** Comparisons in predictive performance of the ML models using a random forest algorithm for pCR prediction with the included PET data.

ML Model	AUC	ACC	F1	Precision	Recall
PET1	0.934 <sup>*,†</sup>	0.827 <sup>*,†</sup>	0.853 <sup>*,†</sup>	0.802 <sup>*,†</sup>	0.912 <sup>†</sup>
PET2	0.975 <sup>*</sup>	0.902 <sup>*,‡</sup>	0.912 <sup>*,‡</sup>	0.905 <sup>*,‡</sup>	0.920
PET3	0.977 <sup>†</sup>	0.934 <sup>†,‡</sup>	0.940 <sup>†,‡</sup>	0.937 <sup>†,‡</sup>	0.944 <sup>†</sup>

AUC, area under curve; ACC, accuracy; PET3, combining PET1 and PET2; \*, †, ‡,  $p < 0.05$ .

**Figure 3.** Comparisons of the ROC curves of the ML models according to the included PET data. It showed that the AUC of ML using PET/CT data obtained after neoadjuvant CCRT was significantly higher than that of using only baseline PET/CT data ( $p < 0.05$ ).

Additionally, we investigated the predictive results from the ML model using four feature subsets with the top 10, 20, 30, and all features from PET3 (Supplemental Table S2 and Supplemental Figure S1). The ML model outperformed other methods when all features were selected (AUC = 0.977, ACC = 0.934, F1 = 0.940, Precision = 0.937, Recall = 0.944).

### 3.3. Predictive Performances of Conventional PET Parameters and Physicians for pCR Prediction

In conventional PET parameters, the SUVmax, SUVmean, MTV, and TLG of PET1 and the SUVmax and SUVmean of PET2 were significantly associated with the pCR ( $p < 0.05$ ). The difference between PET1 and PET2 of the SUVmax ( $p < 0.001$ ), SUVmean ( $p < 0.001$ ), MTV ( $p = 0.003$ ), and TLG ( $p < 0.001$ ) were also significantly associated with the pCR. In contrast, the MTV and TLG of PET2 were not statistically associated with the pCR (Table 3).

**Table 3.** Comparisons in conventional PET parameters according to the presence of pCR.

			Pathologic Response		<i>p</i> -Value
			pCR	Non-pCR	
PET1	SUVmax	Median	13.59	11.58	0.029 *
		IQR	10.01–17.47	8.35–15.53	
	SUVmean	Median	5.91	5.28	0.037 *
		IQR	4.86–7.48	3.97–6.69	
	MTV (cm <sup>3</sup> )	Median	42.96	21.13	0.003 *
		IQR	16.02–74.89	7.38–47.48	
TLG	Median	223.26	113.63	0.001 *	
	IQR	96.29–436.26	30.77–279.36		
PET2	SUVmax	Median	3.17	4.57	<0.001 *
		IQR	2.22–4.13	2.92–6.98	
	SUVmean	Median	1.69	2.35	<0.001 *
		IQR	1.43–2.15	1.74–3.33	
	MTV (cm <sup>3</sup> )	Median	10.40	8.71	0.327
		IQR	3.64–27.11	3.64–19.46	
TLG	Median	19.42	22.00	0.475	
	IQR	6.32–47.35	8.61–56.52		
Delta PET parameters (%)	dSUVmax	Median	74.68	58.14	<0.001 *
		IQR	64.25–84.25	36.07–74.20	
	dSUVmean	Median	70.17	50.79	<0.001 *
		IQR	54.34–78.57	31.58–66.28	
	dMTV (cm <sup>3</sup> )	Median	68.63	48.18	0.003 *
		IQR	42.81–82.49	14.76–71.75	
dTLG	Median	89.52	73.68	<0.001 *	
	IQR	79.40–95.47	50.80–88.83		

pCR, pathologic complete response; PET, positron emission tomography; SUV, standard uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis; IQR, interquartile range; \*, *p* < 0.05.

The optimal cutoff values that allowed significant association with the pCR were PET1-SUVmax = 13.15, PET1-SUVmean = 4.70, PET1-MTV = 41.11, PET1-TLG = 142.97, PET2-SUVmax = 3.97, PET2-SUVmean = 1.83, dSUVmax = 56.5%, dSUVmean = 43.9%, dMTV = 55.4%, and dTLG = 86.2%. Using these cutoff values, the predictive performance of the PET parameters are listed in Table 4. The predictive performance of the physicians based on their diagnostic result are also presented in Table 4.

**Table 4.** Comparisons of predictive performance from conventional PET parameters, from physicians and from the ML model.

	Cutoff	AUC	Sen (%)	Spe (%)	PPV (%)	NPV (%)	ACC (%)
PET1-SUVmax	>13.15	0.592	57.4	61.7	17.7	90.9	61.2
PET1-SUVmean	>4.70	0.588	79.6	39.1	15.8	93.0	44.2
PET1-MTV (cm <sup>3</sup> )	>41.11	0.627	53.7	70.2	20.6	91.3	68.1
PET1-TLG	>142.97	0.635	68.5	57.1	18.9	92.7	59.1
PET2-SUVmax	≤3.97	0.687	74.1	58.8	20.5	94.0	60.7
PET2-SUVmean	≤1.83	0.726	66.7	71.5	25.2	93.7	70.9
dSUVmax	>56.5%	0.737	88.9	48.7	19.9	96.8	53.7
dSUVmean	>43.9%	0.745	94.4	42.8	19.2	98.2	49.3
dMTV (cm <sup>3</sup> )	>55.4%	0.625	68.5	56.6	18.5	92.6	58.1
dTLG	>86.2%	0.703	68.5	69.1	24.2	93.9	69.1
Physicians			33.9	86.4	29.2	90.8	80.5
ML model		0.977	94.4	92.2	93.7	93.1	93.4

AUC, area under curve; Sen, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value; ACC, accuracy.

### 3.4. Comparisons of the ML Model with Conventional PET Parameters and Physicians

A comparison of the predictive performances between conventional PET parameters, physicians, and the ML model are shown in Table 4. First, the performance of the ML



model for pCR prediction was compared with those of conventional PET parameters by analyzing the AUCs. The ML model revealed higher AUC values than all of the single PET parameters ( $p < 0.001$ ). When the pCR was predicted with the conventional single PET parameter, the AUC was only 0.588 to 0.745. By applying the ML model using variable radiomic features, however, the AUC improved to 0.977. In terms of predictive performance, the ML model showed significantly higher performance in Spe, PPV, and ACC than was achieved with any of the conventional PET parameters ( $p < 0.001$ ). When comparing the predictive performances of physicians and of the ML model, the ACC of the ML model was significantly higher than that of physicians (93.4 vs. 80.5%,  $p < 0.001$ ). Not only ACC, but also Sen, Spe, and PPV showed that the ML model significantly increased the results of physicians (94.4 vs. 33.9%,  $p < 0.001$ ; 92.2 vs. 86.4%,  $p = 0.001$ ; 93.7 vs. 29.2%,  $p < 0.001$ ; respectively). NPV was the only case where there was no significant difference between the ML model and prediction by physicians (93.1 vs. 90.8%,  $p = 0.155$ ).

#### 4. Discussion

We have demonstrated that the ML model using an RF algorithm could be robust and useful in determining the pCR following neoadjuvant CCRT by radiomic features of  $^{18}\text{F}$ -FDG PET/CT. Although several studies evaluating ML for treatment response have been published recently [28–31], they mainly conducted research with multiparametric MRI features and not with  $^{18}\text{F}$ -FDG PET/CT. Only a few studies have used  $^{18}\text{F}$ -FDG PET/CT features to assess neoadjuvant treatment response in breast and rectal cancer using ML models [26,27]. To the best of our knowledge, this is the first study to predict the response to neoadjuvant CCRT in patients with NSCLC using an ML model.

The response to neoadjuvant CCRT is critical because it affects postoperative treatment and individual prognosis. Furthermore, the correct prediction of the pCR can determine which patients will require more or less aggressive adjuvant treatment to reduce the risk of complications. Despite improvements in therapeutic modalities of neoadjuvant CCRT, the pCR rate still remains with a variety of outcomes. The gold standard for assessing the pCR is based on postoperative histopathologic findings, which could be inefficient to implement in all patients with advanced NSCLC. Therefore, it is necessary to develop a method of improving the predictive significance of non-invasive imaging modalities for establishing a personalized therapeutic strategy.

Radiomics is an emerging field where various imaging modalities are performed to extract features that may reflect changes in human tissues at the cellular levels and estimate detailed information on tumor biology and microenvironment in nuclear medicine [32,33]. The radiomic features delineated on PET/CT images can represent tumor heterogeneity including fractal dimension, tumor shape, and proliferation [34]. In our experiments, voxel statistics of radiomic features were highly ranked in the prediction for the pCR, followed by texture spectrum and co-occurrence matrix. Although there are differences in the feature importance of many radiomic variables, the ML model using them demonstrated better predictive performance for the pCR than the single conventional PET/CT parameters. Conventional PET parameters and their changes in FDG uptake before and after CCRT have been previously evaluated in determining the treatment response in patients with NSCLC [11]. We also performed these analyses; however, the ACC of the predictive performance using them was only shown to be 44.2–70.9%. Therefore, it seemed unfavorable to evaluate the predictive performance using single PET parameters even though they were statistically significantly correlated with the pCR.

The ML model significantly outperformed the physicians in terms of Sen, Spe, PPV, and ACC. The outcomes of conducting the ML model with PET2 data revealed higher predictive performance than those of the ML model with PET1 data. It appears that radiomic features obtained from PET/CT after neoadjuvant CCRT have more relevant clinical value in the prediction of the pCR. Compared to the results of the ML model with only the variables from each time of PET/CT images, the predictive performance also increased by inputting all variables from both PET1 and PET2. We assumed that the improvement in

performance is probably because of the feature importance for predicting the pCR, which is somewhat different between radiomics of PET1 and PET2. If more significant variables were input into the ML model, the predictive performance may be further improved. The PET-based radiomics can provide the potential to characterize intratumoral heterogeneity indicating resistance to neoadjuvant CCRT. Therefore, it is clinically important to evaluate treatment response not only to obtain baseline PET/CT images but also to examine PET/CT after neoadjuvant CCRT. As the current study demonstrated, the use of ML with radiomics features could be predictive of treatment response and thus help to select a more aggressive treatment for those with high-risk factors after curative surgery in patients with stage III NSCLC.

This study had several limitations. First, this study was conducted in a retrospective manner with a limited sample size from a single center. Because radiomic features can be highly dependent on reconstruction methods and imaging parameters [35], it is planned to obtain a prospective multicenter trial to be more generalizable in the future. Second, the study population was composed of patients with different therapeutic schemes. Although we addressed a homogeneous population of patients with stage III NSCLC, it is also needed to select patients with a more uniform therapeutic modality based on the consistent guideline. Third, various pulmonary side effects can arise after radiotherapy, such as pneumonitis or fibrosis, which may challenge the response assessment, although we tried our best to exclude the possibility of treatment-induced inflammatory changes based on the relative intensity and distribution of FDG uptake in the lung parenchyma and automatically generated tumor VOI [36]. Finally, although the proposed ML model was analyzed using a 10-fold cross-validation for minimizing overfitting instead of splitting the dataset into training and test sets, external validation using an independent dataset is necessary to verify the clinical significance using a larger cohort.

## 5. Conclusions

In conclusion, the developed ML model using an RF algorithm and <sup>18</sup>F-FDG PET/CT radiomics features was useful for predicting the pCR after neoadjuvant CCRT in NSCLC. The predictions of the ML model had higher accuracy than predictions from conventional PET parameters and from physicians. The ML model using radiomics features can be used to facilitate the preoperative individualized prediction for the pCR. Our findings further highlight the potential, non-invasive, and effective clinical significance of an ML model to predict the pCR in patients with stage III NSCLC who had received neoadjuvant CCRT followed by surgery.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14081987/s1>, Figure S1: Comparison of ROC curves from random forest according to ranking-based feature selection; Table S1: List of quantitative PET-based radiomic features from CGITA; Table S2: Predictive performance of random forest according to ranking-based feature selection

**Author Contributions:** All authors contributed equally to this work. Conceptualization, J.Y. and J.Y.C.; data curation, J.Y.; formal analysis, J.Y., J.L. and J.Y.C.; funding acquisition, J.Y. and J.Y.C.; investigation, J.Y. and J.L.; methodology, J.Y., J.L. and J.Y.C.; project administration, J.Y.C.; resources, M.C., S.-K.W., M.-J.A., H.R.P., Y.S.C. and J.H.H.; software, J.Y. and J.L.; supervision, J.Y.C.; validation, J.Y. and J.Y.C.; visualization, J.Y. and J.L.; writing—original draft preparation, J.Y.; writing—review and editing, M.C., S.-K.W., M.-J.A., H.R.P., Y.S.C., J.H.H. and J.Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. NRF-2020M2D9A1094072), Future Medicine 20\*30 Project of the Samsung Medical Center (#SMO1220071), and VHS Medical Center Research Grant (No. VHSMC 22001).

**Institutional Review Board Statement:** This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the Samsung Medical Center (IRB No. 2020-09-185).

**Informed Consent Statement:** Patient consent was waived due to the retrospective design of this study.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from the Samsung Medical Center and are available from the corresponding author with the permission of the Samsung Medical Center.

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

## References

1. DeSantis, C.E.; Lin, C.C.; Mariotto, A.B.; Siegel, R.L.; Stein, K.D.; Kramer, J.L.; Alteri, R.; Robbins, A.S.; Jemal, A. Cancer treatment and survivorship statistics, 2014. *CA Cancer J. Clin.* **2014**, *64*, 252–271. [\[CrossRef\]](#)
2. Arbour, K.C.; Riely, G.J. Systemic therapy for locally advanced and metastatic non-small cell lung cancer: A review. *JAMA* **2019**, *322*, 764–774. [\[CrossRef\]](#)
3. Kim, H.K.; Cho, J.H.; Choi, Y.S.; Zo, J.I.; Shim, Y.M.; Park, K.; Ahn, M.-J.; Ahn, Y.C.; Kim, K.; Kim, J. Outcomes of neoadjuvant concurrent chemoradiotherapy followed by surgery for non-small-cell lung cancer with N2 disease. *Lung Cancer* **2016**, *96*, 56–62. [\[CrossRef\]](#)
4. Hyun, S.H.; Ahn, H.K.; Ahn, M.J.; Ahn, Y.C.; Kim, J.; Shim, Y.M.; Choi, J.Y. Volume-based assessment with 18F-FDG PET/CT improves outcome prediction for patients with stage IIIA-N2 non-small cell lung cancer. *AJR Am. J. Roentgenol.* **2015**, *205*, 623–628. [\[CrossRef\]](#)
5. Schreiner, W.; Gavrychenkova, S.; Dudek, W.; Rieker, R.J.; Lettmaier, S.; Fietkau, R.; Sirbu, H. Pathologic complete response after induction therapy—the role of surgery in stage IIIA/B locally advanced non-small cell lung cancer. *J. Thorac. Dis.* **2018**, *10*, 2795–2803. [\[CrossRef\]](#)
6. Tenahashi, M.; Niwa, H.; Yukiue, H.; Suzuki, E.; Yoshii, N.; Watanabe, T.; Kaminuma, Y.; Chiba, K.; Tsuchida, H.; Yobita, S. Feasibility and prognostic benefit of induction chemoradiotherapy followed by surgery in patients with locally advanced non-small cell lung cancer. *J. Thorac. Dis.* **2020**, *12*, 2644–2653. [\[CrossRef\]](#)
7. Kim, A.W.; Liptay, M.J.; Bonomi, P.; Kim, A.W.; Liptay, M.J.; Bonomi, P.; Warren, W.H.; Basu, S.; Farlow, E.C.; Faber, L.P. Neoadjuvant chemoradiation for clinically advanced non-small-cell lung cancer: An analysis of 233 patients. *Ann. Thorac. Surg.* **2011**, *92*, 233–241, discussion 241–243. [\[CrossRef\]](#)
8. Pottgen, C.; Eberhardt, W.; Graupner, B.; Theegarten, D.; Gauler, T.; Freitag, L.; Jawad, J.A.; Wohlschlaeger, J.; Welter, S.; Stamatis, G.; et al. Accelerated hyperfractionated radiotherapy within trimodality therapy concepts for stage IIIA/B non-small-cell lung cancer: Markedly higher rates of pathologic complete remissions than with conventional fractionation. *Eur. J. Cancer* **2013**, *49*, 2107–2115. [\[CrossRef\]](#)
9. D’Angelillo, R.M.; Trodella, L.; Ciresa, M.; Cellini, F.; Fiore, M.; Greco, C.; Pompeo, E.; Mineo, T.C.; Paleari, L.; Granone, P.; et al. Multimodality treatment of stage III non-small cell lung cancer: Analysis of a phase III trial using preoperative cisplatin and gemcitabine with concurrent radiotherapy. *J. Thorac. Oncol.* **2009**, *4*, 1517–1523. [\[CrossRef\]](#)
10. Stupp, R.; Mayer, M.; Kann, R.; Weder, W.; Zouhair, A.; Betticher, D.C.; Pless, M. Neoadjuvant chemotherapy and radiotherapy followed by surgery in selected patients with stage IIIB non-small-cell lung cancer: A multicentre phase III trial. *Lancet Oncol.* **2009**, *10*, 785–793. [\[CrossRef\]](#)
11. Cremonesi, M.; Gilardi, L.; Ferrari, M.E.; Piperno, G.; Travaini, L.L.; Timmerman, R.; Botta, F.; Baroni, G.; Grana, C.M.; Ronchi, S.; et al. Role of interim 18F-FDG PET/CT for the early prediction of clinical outcomes of non-small cell lung cancer (NSCLC) during radiotherapy or chemo-radiotherapy. A systematic review. *Eur. J. Nucl. Med. Mol. Imaging* **2017**, *44*, 1915–1927. [\[CrossRef\]](#)
12. Roengvoraphoj, O.; Wijaya, C.; Eze, C.; Li, M.; Dantes, M.; Taugner, J.; Tufman, A.; Huber, R.M.; Belka, C.; Manapov, F. Analysis of primary tumor metabolic volume during chemoradiotherapy in locally advanced non-small cell lung cancer. *Strahlenther. Onkol.* **2018**, *194*, 107–115. [\[CrossRef\]](#)
13. Pöttgen, C.; Levegrün, S.; Theegarten, D.; Marnitz, S.; Grehl, S.; Pink, R.; Eberhardt, W.; Stamatis, G.; Gauler, T.; Antoch, G.; et al. Value of <sup>18</sup>F-fluoro-2-deoxy-D-glucose-positron emission tomography/computed tomography in non-small-cell lung cancer for prediction of pathologic response and times to relapse after neoadjuvant chemoradiotherapy. *Clin. Cancer Res.* **2006**, *12*, 97–106. [\[CrossRef\]](#)
14. Cerfolio, R.J.; Bryant, A.S.; Winokur, T.S.; Ohja, B.; Bartolucci, A.A. Repeat FDG-PET after neoadjuvant therapy is a predictor of pathologic response in patients with non-small cell lung cancer. *Ann. Thorac. Surg.* **2004**, *78*, 1903–1909. [\[CrossRef\]](#)
15. 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.* **2009**, *50*, 1225–150S. [\[CrossRef\]](#)
16. Joo Hyun, O.; Lodge, M.A.; Wahl, R.L. Practical PERCIST: A simplified guide to PET response criteria in solid tumors 1.0. *Radiology* **2016**, *280*, 576–584.

17. De Ruyscher, D.; Houben, A.; Aerts, H.J.; Dehing, C.; Wanders, R.; Öllers, M.; Lambin, P. Increased (18)F-deoxyglucose uptake in the lung during the first weeks of radiotherapy is correlated with subsequent radiation-induced lung toxicity (RILT): A prospective pilot study. *Radiother. Oncol.* **2009**, *91*, 415–420. [[CrossRef](#)]
18. Szyszko, T.A.; Yip, C.; Szlosarek, P.; Goh, V.; Cook, G.J. The role of new PET tracers for lung cancer. *Lung Cancer* **2019**, *94*, 7–14. [[CrossRef](#)] [[PubMed](#)]
19. Yoo, J.; Cheon, M.; Park, Y.J.; Hyun, S.H.; Zo, J.I.; Um, S.W.; Choi, J.Y. Machine learning-based diagnostic method of pre-therapeutic <sup>18</sup>F-FDG PET/CT for evaluating mediastinal lymph nodes in non-small cell lung cancer. *Eur. Radiol.* **2021**, *31*, 4184–4194. [[CrossRef](#)]
20. Rami-Porta, R.; Crowley, J.J.; Goldstraw, P. The revised TNM staging system for lung cancer. *Ann. Thorac. Cardiovasc. Surg.* **2009**, *15*, 4–9.
21. Shin, S.; Kim, H.K.; Cho, J.H.; Choi, Y.S.; Kim, K.; Kim, J.; Zo, J.I.; Sun, J.; Ahn, M.; Park, K.; et al. Adjuvant therapy in stage IIIA-N2 non-small cell lung cancer after neoadjuvant concurrent chemotherapy followed surgery. *J. Thorac. Dis.* **2020**, *12*, 2602–2613. [[CrossRef](#)] [[PubMed](#)]
22. Cottrell, T.R.; Thompson, E.D.; Forde, P.M.; Stein, J.E.; Duffield, A.S.; Anagnostou, V.; Rekhman, N.; Anders, R.A.; Cuda, J.D.; Illei, P.B.; et al. Pathologic features of response to neoadjuvant anti-PD-1 in resected non-small-cell lung carcinoma: A proposal for quantitative immune-related pathologic response criteria (irPRC). *Ann. Oncol.* **2018**, *29*, 1853–1860. [[CrossRef](#)] [[PubMed](#)]
23. Mouillet, G.; Monnet, E.; Milleron, B.; Puyraveau, M.; Quoix, E.; David, P.; Ducloné, A.; Molinier, O.; Zalcman, G.; Depierre, A.; et al. Pathologic complete response to preoperative chemotherapy predicts cure in early-stage non-small-cell lung cancer: Combined analysis of two IFCT randomized trials. *J. Thorac. Oncol.* **2012**, *7*, 841–849. [[CrossRef](#)] [[PubMed](#)]
24. Li, P.; Wang, X.; Xu, C.; Liu, C.; Zheng, C.; Fulham, M.J.; Feng, D.; Wang, L.; Song, S.; Huang, G. (18)F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 1116–1126. [[CrossRef](#)] [[PubMed](#)]
25. Shen, W.C.; Chen, S.W.; Wu, K.C.; Lee, P.Y.; Feng, C.L.; Hsieh, T.C.; Yen, K.; Kao, C. Predicting pathological complete response in rectal cancer after chemoradiotherapy with a random forest using 18F-fluorodeoxyglucose positron emission tomography and computed tomography radiomics. *Ann. Transl. Med.* **2020**, *8*, 207. [[CrossRef](#)] [[PubMed](#)]
26. Tahmassebi, A.; Wengert, G.J.; Helbich, T.H.; Bago-Horvath, Z.; Alaei, S.; Bartsch, R.; Dubsky, P.; Baltzer, P.; Clauser, P.; Kapetas, P.; et al. Impact of machine learning with parametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients. *Investig. Radiol.* **2019**, *54*, 110–117. [[CrossRef](#)]
27. Yakar, M.; Etiz, D.; Metintas, M.; Ak, G.; Celik, O. Prediction of radiation pneumonitis with machine learning in stage III lung cancer: A pilot study. *Technol. Cancer Res. Treat.* **2021**, *20*, 15330338211016373. [[CrossRef](#)]
28. Meti, N.; Saednia, K.; Lagree, A.; Tabbarah, S.; Mohebpour, M.; Kiss, A.; Lu, F.; Slodkowska, E.; Gandhi, S.; Jerzak, K.J.; et al. Machine learning frameworks to predict neoadjuvant chemotherapy response in breast cancer using clinical and pathological features. *JCO Clin. Cancer Inform.* **2021**, *5*, 66–80. [[CrossRef](#)]
29. Lo Gullo, R.; Eskreis-Winkler, S.; Morris, E.A.; Pinker, K. Machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy. *Breast* **2020**, *49*, 115–122. [[CrossRef](#)]
30. Huang, C.; Huang, M.; Huang, C.; Tsai, H.; Su, W.; Chang, W.; Wang, J.; Shi, H. Machine learning for predicting pathological complete response in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy. *Sci. Rep.* **2020**, *10*, 12555. [[CrossRef](#)]
31. Eun, N.L.; Kang, D.; Son, E.J.; Park, J.S.; Youk, J.H.; Kim, J.A.; Gweon, H.M. Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer. *Radiology* **2020**, *294*, 31–41. [[CrossRef](#)] [[PubMed](#)]
32. Antunovic, L.; de Sanctis, R.; Cozzi, L.; Kirienko, M.; Sagona, A.; Torrasi, R.; Tinterri, C.; Santoro, A.; Chiti, A.; Zelic, R.; et al. PET/CT radiomics in breast cancer: Promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 1468–1477. [[CrossRef](#)] [[PubMed](#)]
33. Ha, S.; Park, S.; Bang, J.I.; Kim, E.K.; Lee, H.Y. Metabolic radiomics for pretreatment (18)F-FDG PET/CT to characterize locally advanced breast cancer: Histopathologic characteristics, response to neoadjuvant chemotherapy, and prognosis. *Sci. Rep.* **2017**, *7*, 1556. [[CrossRef](#)] [[PubMed](#)]
34. Hoffmann, B.; Frenzel, T.; Schmitz, R.; Schumacher, U.; Wedemann, G. Modeling growth of tumors and their spreading behavior using mathematical functions. *Methods Mol. Biol.* **2019**, *1878*, 263–277. [[PubMed](#)]
35. Sollini, M.; Cozzi, L.; Antunovic, L.; Chiti, A.; Kirienko, M. PET radiomics in NSCLC: State of the art and a proposal for harmonization of methodology. *Sci. Rep.* **2017**, *7*, 358. [[CrossRef](#)] [[PubMed](#)]
36. Iravani, A.; Turgeon, G.A.; Akhurst, T.; Callahan, J.W.; Bressel, M.; Everitt, S.J.; Siva, S.; Hofman, M.S.; Hicks, R.J.; Ball, D.L.; et al. PET-detected pneumonitis following curative-intent chemoradiation in non-small cell lung cancer (NSCLC): Recognizing patterns and assessing the impact on the predictive ability of FDG-PET/CT response assessment. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 1869–1877. [[CrossRef](#)]