



Multicentric development and evaluation of [¹⁸F]FDG PET/CT and CT radiomic models to predict regional and/or distant recurrence in early-stage non-small cell lung cancer treated by stereotactic body radiation therapy

François Lucia^{1,2,3,4} · Thomas Louis³ · François Cousin³ · Vincent Bourbonne^{1,2} · Dimitris Visvikis² · Carole Mievis⁵ · Nicolas Jansen⁵ · Bernard Duysinx⁶ · Romain Le Pennec^{7,8} · Malik Nebbache¹ · Martin Rehn¹ · Mohamed Hamya¹ · Margaux Geier⁹ · Pierre-Yves Salaun^{7,8} · Ulrike Schick^{1,2} · Mathieu Hatt² · Philippe Coucke⁵ · Roland Hustinx^{3,10} · Pierre Lovinfosse³

Received: 29 August 2023 / Accepted: 3 November 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Purpose To develop machine learning models to predict regional and/or distant recurrence in patients with early-stage non-small cell lung cancer (ES-NSCLC) after stereotactic body radiation therapy (SBRT) using [¹⁸F]FDG PET/CT and CT radiomics combined with clinical and dosimetric parameters.

Methods We retrospectively collected 464 patients (60% for training and 40% for testing) from University Hospital of Liège and 63 patients from University Hospital of Brest (external testing set) with ES-NSCLC treated with SBRT between 2010 and 2020 and who had undergone pretreatment [¹⁸F]FDG PET/CT and planning CT. Radiomic features were extracted using the PyRadiomics toolbox®. The ComBat harmonization method was applied to reduce the batch effect between centers. Clinical, radiomic, and combined models were trained and tested using a neural network approach to predict regional and/or distant recurrence.

Results In the training ($n=273$) and testing sets ($n=191$ and $n=63$), the clinical model achieved moderate performances to predict regional and/or distant recurrence with *C*-statistics from 0.53 to 0.59 (95% CI, 0.41, 0.67). The radiomic (original_firstorder_Entropy, original_gldm_LowGrayLevelEmphasis and original_gldm_DifferenceAverage) model achieved higher predictive ability in the training set and kept the same performance in the testing sets, with *C*-statistics from 0.70 to 0.78 (95% CI, 0.63, 0.88) while the combined model performs moderately well with *C*-statistics from 0.50 to 0.62 (95% CI, 0.37, 0.69).

Conclusion Radiomic features extracted from pre-SBRT analog and digital [¹⁸F]FDG PET/CT outperform clinical parameters in the prediction of regional and/or distant recurrence and to discuss an adjuvant systemic treatment in ES-NSCLC. Prospective validation of our models should now be carried out.

Keywords Non-small cell lung cancer · Stereotactic body radiation therapy · Radiomics · Machine learning · [¹⁸F]FDG PET/CT

Introduction

Stereotactic body radiation therapy (SBRT) has become a therapeutic standard for inoperable early-stage non-small cell lung cancer (ES-NSCLC). Numerous studies have shown very excellent local control and minimal toxicity [1–3]. Although this new technique has been broadly recognized for its remarkably high rate of local control, the dominant risk of recurrence is the occurrence of distant

François Lucia, Thomas Louis, Roland Hustinx, and Pierre Lovinfosse equally contributed.

Extended author information available on the last page of the article

metastases [4]. Moreover, dosimetric parameters seem to have an impact on local control (LC) but not on the risk of regional or distant relapse [5].

Ongoing studies are evaluating the benefit of concomitant treatment, including immunotherapy, to SBRT (NCT03924869, NCT03050554, NCT03833154), and one study showed an improvement in recurrence-free survival with immunotherapy [6]. Thus, it would be interesting to have non-invasive tools to select patients at high risk of regional or distant relapse who could benefit from concomitant or adjuvant systemic therapy to avoid their side effects [6]. Gao et al. developed a multicenter model based on clinical parameters with internal and external validation to predict distant metastatic spread [7]. The model identified a high-risk patient subgroup who had greater rates of distant metastases. However, this model had a medium sensitivity and was not statistically significant in the testing sets [7]. The authors highlighted the potential of radiomics to improve results.

Fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F]FDG PET/CT) is a molecular imaging technique combining metabolic and functional assessment, enabling improved diagnostic accuracy and initial staging and restaging of lung cancer, as well as optimization of treatment and follow-up of therapeutic response [8]. Regarding the prognostic impact of ^{18}F]FDG PET/CT, results are divergent for the standardized uptake (SUV) and metabolic tumor volume on outcomes [9–13].

It has been shown that tumor heterogeneity could be studied by quantitative analysis of medical images, including ^{18}F]FDG PET/CT and CT. These complementary analyses have enabled the development of predictive and prognostic models for ES-NSCLC [14–18]. However, to the best of our knowledge, no radiomic studies have included a complete dosimetric analysis which may be an important bias in predicting the risk of recurrence. Machine learning (ML) methods, in particular artificial neural networks (ANN) [19, 20], allow to build predictive models by combining parameters using a flexible non-linear relationship [21].

The aim of this study was to develop ML models to predict the risk of regional and/or distant recurrence in patients with ES-NSCLC treated by SBRT using ^{18}F]FDG PET/CT and/or CT radiomic features from the primary tumor volume, combined (or not) with clinical parameters.

Methods

Patients' information

Patients with histologically or clinically diagnosed ES-NSCLC T1 (<3 cm) and T2 (3–5 cm) without lymph node involvement treated with SBRT between April 2010 and

December 2020 from two institutions (University Hospital (CHU) of Liège in Belgium and CHU of Brest in France) were retrospectively considered. Clinical and pathological data included age, date of diagnosis, gender, WHO performance status (PS), clinical T stage, tumor size as measured on CT (according to the report of the radiologist), histology, mutational status, tumor size, smoking history, history of lung surgery or radiation therapy, and peripheral versus central (defined by a tumor located within 2 cm of the proximal bronchial) tumor location (Table S1). We also collected detailed dosimetric parameters. Exclusion criteria were locally advanced NSCLC, metastatic tumors, cancers other than non-small cell histology, a history of lung cancer within the last 5 years, incomplete course of SBRT, and a concomitant or adjuvant systemic treatment. For patients who could not undergo or refused biopsy, the clinical diagnosis of NSCLC was validated by multidisciplinary tumor board, including a clinical lung cancer pulmonologist, radiologist, nuclear medicine physician, and radiation oncologist, on the basis of strong imaging suspicion that showed a high uptake on ^{18}F]FDG PET/CT and an increase in two consecutive CT scans acquired 3 months apart. Clinical staging of the lung cancer was performed according to the 8th Union for International Cancer Control TNM staging system using CT, brain imaging (magnetic resonance imaging (MRI) or CT), and ^{18}F]FDG PET/CT. Of the 597 eligible patients, 70 patients were not included because ^{18}F]FDG PET/CT was not available ($n=50$) or because of poor quality of PET/CT or CT imaging ($n=20$) (supplemental data D, figure S1). Finally, a total of 527 patients were included consisting of 63 patients from the CHU of Brest and 464 at the CHU of Liège. Regarding the 63 patients from the CHU of Brest, 21 who had an analog PET/CT (2016–2018) and 42 who had a digital PET/CT (2019–2020).

The study was approved by both institutional ethical committees.

Treatment

Detailed treatment protocols are available in the supplemental data A.

At the CHU of Liège, SBRT was administered using a dedicated machine (CyberKnife®, Accuray). For each patient, the most suitable respiratory gating method was chosen [22]. Prescription doses at the 80% isodose line of PTV ranged from 40 to 60 Gy in 3 to 5 fractions.

At the CHU of Brest, a respiratory-sorted 4-dimensional computed tomography (4DCT) dataset was generated using the planning CT (Siemens, Somatom) coupled with Varian real-time position management (RPM) gating system. The prescribing goal was that 99% of the planning target volume (PTV) should receive at least 99% of the dose, with a maximum dose to the PTV of up to 140% of the prescribed dose.

Prescribed doses ranged from 48 to 60 Gy in 3 to 5 fractions. All treatments were delivered by volumetric modulated arc therapy (VMAT).

Endpoints

Regional failure was defined as lymph node metastasis in the hilar, mediastinal, or supraclavicular lymph node stations. Metastatic recurrence was defined as failure in the same lung lobe (farther than 1.5 cm from the primary tumor), in other pulmonary lobes (ipsi- or contralateral lung) or in other organs. These failures had to be confirmed histologically or clinically on the basis of CT and [^{18}F]FDG PET/CT results and validated by a multidisciplinary committee. The differential diagnosis between a recurrence and a second primary lung tumor had to take into account the pathology findings, the interval between the relapse and the primary tumor, and the location of the failure in comparison to the SBRT field [23]. In the case of death, the cause was reported to be cancer-specific or not. Specific death was considered when the patient presented with cancer relapse at the time of death, except for patients with another identified cause of death. All events were measured from the first day of radiotherapy (RT). OS was calculated from the first day of RT to the date of death from any cause. Patients alive at the time of analysis were censored upon the last follow-up. Follow-up was calculated using a reverse Kaplan–Meier estimation [24].

PET/CT imaging

PET/CT studies were performed with 3 types of scanners and 3 types of acquisition. At CHU of Liège, studies were acquired using cross-calibrated Philips Gemini TF or BB. At CHU of Brest, studies were performed with a Siemens Biograph mCT (between 2016 and 2018) and with a Siemens digital Biograph Vision 600 (between 2019 and 2020) (supplemental data B, Table S1).

Planning CT imaging

CT studies were performed with 2 types of scanners and 2 types of acquisition. At CHU of Liège, studies were acquired with a Philips Brilliance Big Bore CT (Philips Healthcare). At CHU of Brest, studies were performed with a Siemens Somatom (Siemens Healthcare, Malvern, PA, USA). No contrast-enhancing agent was used (supplemental data B, Table S1).

Tumor volume delineation

[^{18}F]FDG PET exams were imported into a MIM workstation (MIM Software®, Cleveland, OH, USA). Segmentation of the 3D primary tumor volume was semi-automatically

performed by an experienced radiation oncologist (F.L.) using a gradient-based method (PET-Edge®) [25].

The primary tumor was also delineated on (i) the CT of PET/CT and (ii) the planning CT (the expiration CT scan was used at the CHU of Liège and the average was used at the CHU of Brest). Each CT was segmented independently using a previously validated semi-automatic approach exploiting 3D Slicer® and the GrowCut algorithm [26].

Semi-automatic segmentation was performed by one other expert for [^{18}F]FDG PET exams (P.L.) and CT scans (F.C.) blinded to the results of the previous delineation by F.L. in a randomly selected subset ($n = 50$) of the training set to evaluate the inter-reader variability (supplemental data C).

Radiomic feature extraction

One hundred and six radiomic features were extracted from the segmented volumes in each image modality using PyRadiomics v3.0.1 (Boston, MA, USA) [27] and follow the IBSI (imaging biomarker standardization initiative) benchmark [28]. Radiomic features were extracted after $2 \times 2 \times 2 \text{ mm}^3$ spatial resampling of all PET images, and $1 \times 1 \times 1 \text{ mm}^3$ spatial resampling of CT using a cubic spline interpolation. For the calculation of the texture matrix-based features, image intensities were discretized using fixed bin number (FBN, using 64 bins) [28]. As a result, 318 radiomic features (106×3 for PET, CT of PET, and planning CT) were available for each patient. Given the number of PET/CT and CT model configurations and acquisition/reconstruction parameters present in our multicentric database, ComBat a posteriori statistical harmonization method was applied [29]. Then, a normalization with the z -score was carried out because the feature ranges were very different, and high-value features could have a greater impact on classifier performance than low-value features.

Prognosis analysis

Prediction of regional and/or distant recurrence was chosen as the primary endpoint. Prediction of distant recurrence (with or without regional recurrence), isolated regional recurrence, and cancer-specific survival was chosen as the secondary endpoints. In addition to the radiomic features, data also included clinical and dosimetric parameters (Supplemental data B, table S2). Three models have been designed for each endpoint: (I) a standard model (including only clinical and dosimetric data), (II) a radiomic model, and (III) a combined (including clinical, dosimetric, and radiomic features) model. The full dataset of CHU of Liège was separated by random sampling in a training cohort (60%, $n = 273$) and a testing cohort (40%, $n = 191$).

Features were first selected using the Mann–Whitney test after harmonization and normalization. Only statistically

associated features in regard to the outcome were retained. Correlation between the retained features was then assessed with the Spearman correlation coefficient, keeping only the most significant feature in case of a Spearman coefficient ≥ 0.7 . Selected features were then combined using a decremental neural network approach based on each feature's ranking, the ranking being set by the importance of the feature in the proposed model. The neural network was a multilayer perceptron network, embedded in SPSS Modeler V18.3® (IBM, NY, USA). For each model, the least important feature is put apart and the remaining features are being provided for the development of the next model. The chosen model was the one maximizing the mean accuracy based on the 1000 replications of the bootstrapping. Before each model was built, correction for unbalanced data was performed using the Synthetic Minority Over-sampling Technique-Nominal Continuous (SMOTE-NC). Multilayer perceptron is a classifier based on the feedforward artificial neural network (<https://spark.apache.org/docs/latest/ml-classification-regression.html#multilayer-perceptron-classifier>). Multilayer perceptron consists of multiple layers of nodes. Each layer is fully connected to the next layer in the network. Nodes in the input layer represent the input data. All other nodes map inputs to outputs by a linear combination of the inputs with the node's weights w and bias b and applying an activation function. The number of nodes N in the output layer corresponds to the number of classes. Multilayer perceptron employs backpropagation for learning the model. We use the logistic loss function for optimization and L-BFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno) as an optimization routine. Regarding the optimization parameters of our multilayer perceptron network. About the details of each step, the initial tuning parameters were defined as follows: the softmax activation function, an initial lambda of $5 \cdot 10^{-7}$, an initial sigma of $5 \cdot 10^{-5}$, and an interval center of 0 with an interval offset of ± 0.5 . For objectives, we have used and build a new model and enhance model stability (bagging). For basics, we have selected neural network model and multilayer perceptron. For hidden layers, we have selected "automatically comput number of units." For stopping rules, we have chosen "use maximum training time (per component model): 15 min" (this is the default settings). For ensembles, we have chosen voting (default settings). We have used 1000 for the number of component models for boosting and/or bagging: 1000. In the advanced part, we have selected 30 for overfit prevention set and for the replicate results, analyses are replicated with seed 229,176,228 by default. Finally, for missing values in predictors, we have chosen to "delete listwise removes records with missing values on predictors from model building."

The trained models were then evaluated on the testing sets. Statistical analysis was then performed with MedCalc® (version 15.8, MedCalc Software bvba, Ostend, Belgium).

To determine the threshold values for significant parameters (with a p -value ≤ 0.05), the receiver operating characteristic (ROC) curve was used with the Youden index. Models were evaluated on the internal and external testing cohorts using ROC (area under the curve (AUC), sensitivity (Se), specificity (Sp), balanced accuracy (Bacc), and C -statistic [30] as well as F1-Max). Patients have then been dichotomized into low- and high-risk groups. The Kaplan–Meier curves of the models were drawn. The difference between low- and high-risk groups has been tested using the Logrank test. The Harrell concordance index (C -index) was calculated. Decision curves were also used for the models' evaluation.

Results

Among 527 patients who underwent SBRT for ES-NSCLC, regional and distant recurrence was found in 105 (20%) and 143 patients (27%), respectively, without significant differences between cohorts (Supplemental data C, table S2).

Models for the regional and/or distant recurrence

In the training, testing, and external testing sets, 188 (68.9%), 120 (62.8%), and 48 (76.2%) patients had a regional and/or distant recurrence, respectively ($p = 0.11$) (Supplemental data C, table S2).

Thirty PET and CT radiomic features were preselected using the Mann–Whitney test on the training set. At the second step (Spearman rank correlation coefficient), 5 radiomic features from PET imaging only showed intra-correlation levels below 0.7 and were subsequently used to train the models.

The best clinical model combined 3 features (clinical stage, histology, and age). The best radiomic and combined models were based on 3 radiomic features from PET imaging only (original_firstorder_Entropy, original_gldm_LowGrayLevelEmphasis, and original_glcm_DifferenceAverage) and 4 (same radiomic features and clinical stage) variables, respectively, among which the most important one was original_firstorder_Entropy (contribution of 48% and 43%, respectively).

Composition of each model is available in supplemental data B, Table S3.

In the training set ($n = 273$), the clinical model achieved a poor prediction of the risk of regional and/or distant recurrence with an AUC of 0.58 (95% CI, 0.52, 0.64) and a C -statistic of 0.59 (95% CI, 0.53, 0.65) with a cut-off of 70%. The radiomic and combined models achieved higher performance with an AUC of 0.74 (95% CI, 0.69, 0.80) and a C -statistic of 0.71 (95% CI, 0.66, 0.77) with a cut-off of 77% and AUC of 0.67 (95% CI, 0.61, 0.73) and a C -statistic of 0.65 (95% CI, 0.59, 0.71) with a cut-off of 85%, respectively (Table 1). The predictive performance of the clinical, radiomic, and combined models for regional

and/or distant-free survival were significant ($p = 0.0042$, $p < 0.0001$, and $p < 0.0001$, respectively), with a C -index of 0.54 (95% CI, 0.48, 0.61), 0.68 (95% CI, 0.62, 0.74), and 0.59 (95% CI, 0.53, 0.65), respectively (figure S2).

In the testing ($n = 191$) and external testing sets ($n = 63$), the clinical model predicted the risk of regional and/or distant recurrence with lower AUCs of 0.53 (95% CI, 0.46, 0.60) and 0.51 (95% CI, 0.38, 0.64) and C -statistics of 0.53 (95% CI, 0.45, 0.60) and 0.54 (95% CI, 0.41, 0.67), respectively, with a cut-off of 70%. The radiomic model was the most efficient model with AUCs of 0.79 (95% CI, 0.72, 0.84) and 0.83 (95% CI, 0.71, 0.91) and C -statistics of 0.70 (95% CI, 0.63, 0.77) and 0.78 (95% CI, 0.66, 0.88), respectively, with a cut-off of 77%. In comparison, the combined model resulted in lower (although still much higher than clinical only) AUCs of 0.67 (95% CI, 0.60, 0.73) and 0.56 (95% CI, 0.43, 0.68) and C -statistics of 0.62 (95% CI, 0.54, 0.69) and 0.50 (95% CI, 0.37, 0.63), respectively, with a cut-off of 77% (Tables 2 and 3). Only the radiomic model was significantly predictive for regional and/or distant-free survival ($p < 0.05$), with C -indexes of 0.70 (95% CI, 0.64, 0.76) and 0.71 (95% CI, 0.59, 0.82), respectively (Fig. 1).

These models were used to separate patients into two groups: low- and high-risk of regional and/or distant recurrence (Tables 1, 2, and 3; Fig. 2). According to the clinical model, the high-risk patient subgroup had a higher rate of regional and/or distant metastases in the training set (44.4.% [36 of 81] versus 25.5% [49 of 192], $p = 0.0033$), the internal test set (40.6% [28 of 69] versus 35.2% [43 of 122], $p = 0.56$), and the external testing set (29.4% [5 of 17] versus 27.8% [10 of 36], $p = 0.84$). For the radiomic model, the rate of regional and/or distant metastases in the high-risk subgroup was increased in the training set (47.7% [72 of 151] versus 10.7% [13 of 122], $p < 0.0001$), the internal test set (52.5% [62 of 118] versus 12.3% [9 of 73], $p < 0.0001$), and the external testing set (41.7% [15 of 36] versus 0% [0 of 27], $p = 0.0004$). Only the radiomic model was significant in the internal and external testing sets.

Applying both models to the overall cohort, more patients with regional and/or distant recurrence would adequately receive a systemic treatment following the radiomic model ($n = 149$) than the clinical model ($n = 69$). However, unnecessary systemic treatment would be avoided in more patients using the clinical model: 258 vs 200. Conversely, more patients with recurrence would not be treated with the clinical model: 102 vs 22 patients.

Finally, DCA are available in supplemental data D, figure S14.

Models for the distant recurrence, isolated regional recurrence, and cancer-specific survival

These models are available in supplemental data E.

Table 1 Models in Liège training set for regional and/or distant recurrence

Model	AUC	Model cut-off	C-statistic	Se	Sp	Bacc	F1 score	Number of patients					
								Below the cut-off		Above the cut-off			
								Total	Without recurrence	With recurrence	Total	Without recurrence	With recurrence
Clinical	0.58 [0.52–0.64]	70%	0.59 [0.53–0.65]	42.4 [31.7–53.6]	76.1 [69.3–82.0]	59.3 [50.5–67.8]	0.48	192	143	49	81	45	36
Radiomic	0.74 [0.69–0.80]	77%	0.71 [0.66–0.77]	84.7 [75.3–91.6]	58.0 [50.6–65.1]	71.4 [63.0–78.4]	0.61	122	109	13	151	79	72
Combined	0.67 [0.61–0.73]	85%	0.65 [0.59–0.71]	72.9 [62.2–82.0]	56.9 [49.5–64.1]	64.9 [55.9–73.1]	0.54	130	107	23	143	81	62

Data in brackets are 95% CIs; abbreviations: AUC area under the curve, Se sensitivity, Sp specificity, Bacc balanced accuracy

Table 2 Models in Liège testing set for regional and/or distant recurrence

Model	AUC	Model cut-off	C-statistic	Se	Sp	Bacc	F1 score	Number of patients					
								Below the cut-off		Above the cut-off			
								Total	Without recurrence	With recurrence	Total	Without recurrence	With recurrence
Clinical	0.53 [0.46–0.60]	70%	0.53 [0.45–0.60]	39.4 [28.0–51.7]	65.8 [56.6–74.2]	52.6 [42.3–62.9]	0.54	122	79	43	69	41	28
Radiomic	0.79 [0.72–0.84]	77%	0.70 [0.63–0.77]	87.3 [77.3–94.0]	53.3 [44.0–62.5]	70.3 [60.7–78.3]	0.66	73	64	9	118	56	62
Combined	0.67 [0.60–0.73]	85%	0.62 [0.54–0.69]	73.2 [61.4–83.1]	50.0 [40.7–59.3]	61.6 [51.1–71.2]	0.57	79	60	19	112	60	52

Data in brackets are 95% CIs; abbreviations: *AUC* area under the curve, *Se* sensitivity, *Sp* specificity, *Bacc* balanced accuracy

Table 3 Models in Brest external testing set for regional and/or distant recurrence

Model	AUC	Model cut-off	C-statistic	Se	Sp	Bacc	F1 score	Number of patients					
								Below the cut-off		Above the cut-off			
								Total	Without recurrence	With recurrence	Total	Without recurrence	With recurrence
Clinical	0.51 [0.38–0.64]	70%	0.54 [0.41–0.67]	33.3 [11.8–61.6]	75.0 [60.4–86.4]	54.2 [36.1–74.0]	0.39	46	36	10	17	12	5
Radiomic	0.83 [0.71–0.91]	77%	0.78 [0.66–0.88]	100 [78.2–100]	56.2 [41.2–70.5]	78.1 [59.7–85.3]	0.67	27	27	0	36	21	15
Combined	0.56 [0.43–0.68]	85%	0.50 [0.37–0.63]	33.3 [11.8–61.6]	66.7 [51.6–79.6]	50.0 [31.7–70.6]	0.39	42	32	10	21	16	5

Data in brackets are 95% CIs; abbreviations: *AUC* area under the curve, *Se* sensitivity, *Sp* specificity, *Bacc* balanced accuracy

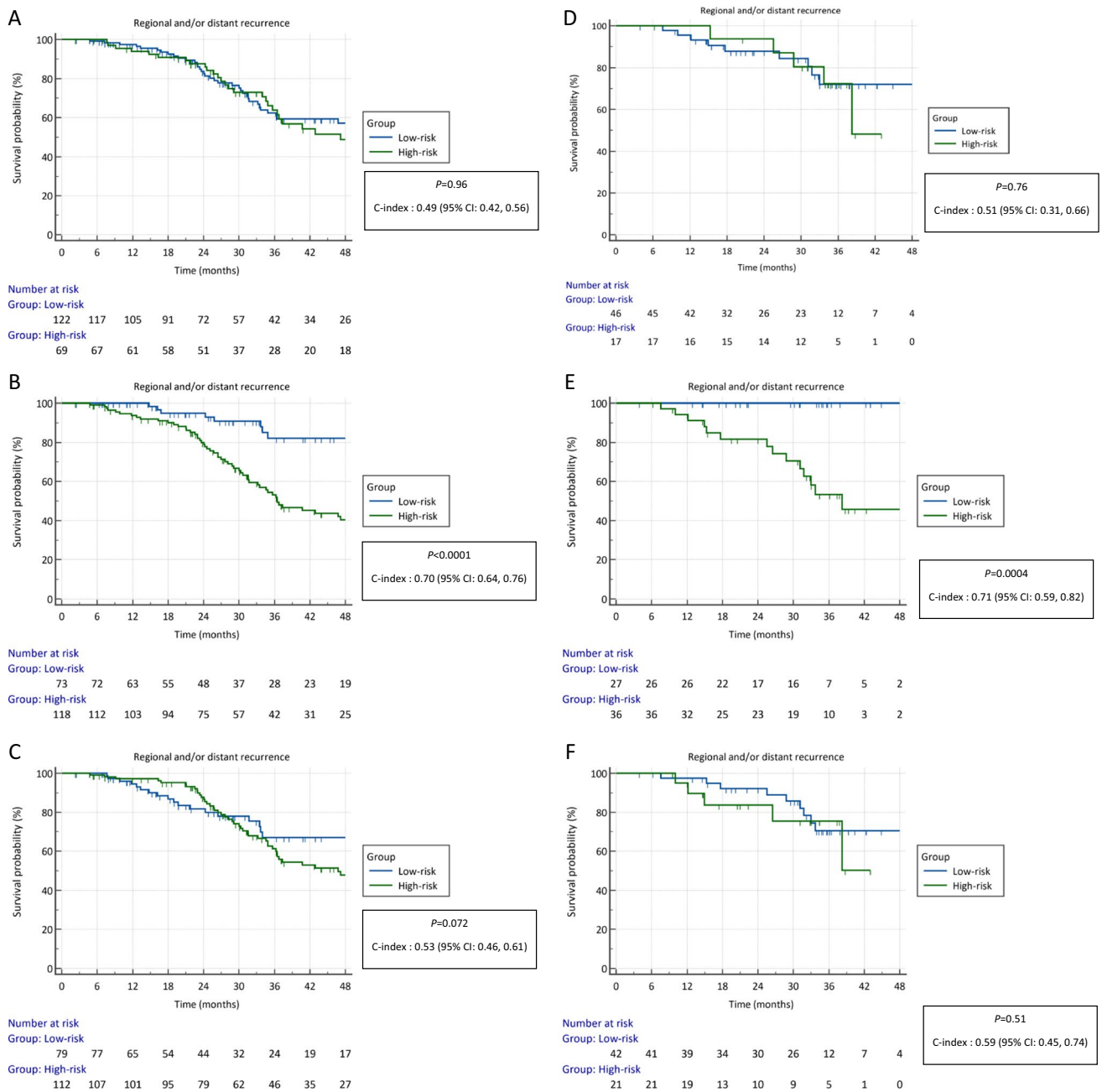


Fig. 1 Kaplan–Meier survival curves for regional and/or distant recurrence in Liège testing set for clinical (A), radiomic (B), and combined (C) models and in Brest external testing set for clinical (D), radiomics (E) and combined (F) models

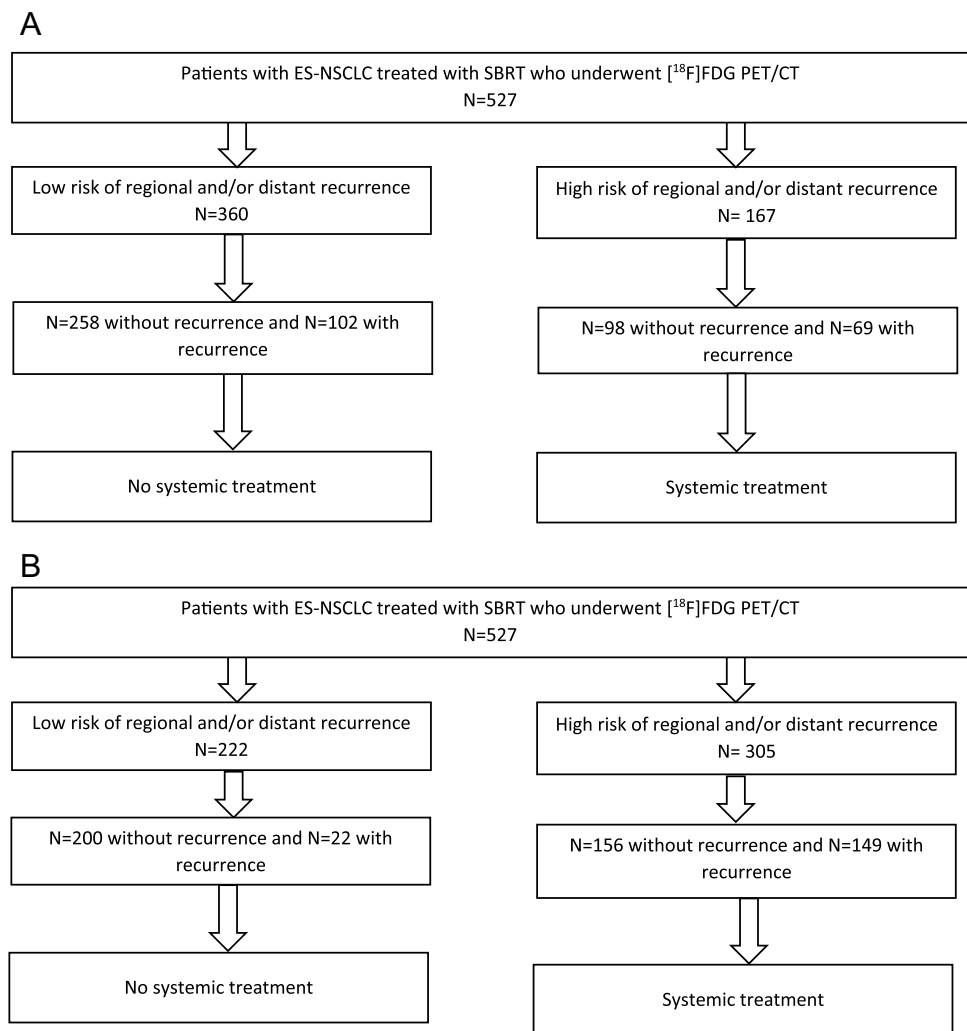
Discussion

In our study, we developed a machine learning radiomic-based model that could be used in clinical practice to predict the risk of regional and/or distant recurrence and potentially guide clinicians in the decision to offer concomitant and/or adjuvant systemic therapy to patients with ES-NSCLC treated by SBRT. This model outperformed

the clinical model with a high sensitivity (from 85 to 100%) and a moderate specificity (from 53 to 58%).

In ES-NSCLC, the dominant pattern of recurrence after local treatment with SBRT is the development of lymph node or distant metastases [4]. Prediction of regional and metastatic recurrence risk before SBRT for ES-NSCLC is a key issue. Systemic therapy, including immunotherapy, could be considered in addition to SBRT in these high-risk patients. A retrospective study showed that some

Fig. 2 Flow diagram of risk-stratification strategy based on pretreatment [^{18}F]FDG PET/CT illustrated in the pooled set for clinical (A) and radiomic models (B). These models separate patients into two groups: low and high risk of regional and/or distant recurrence. High-risk patients could have systemic treatment and low-risk group would be spared of a systemic treatment. Thus, more patients with recurrence would be treated with radiomic model 149 (B) vs 69 (A). However, unnecessary systemic treatment would be avoided in more patients with clinical model 258 (A) vs 200 (B) patients but more patients with recurrence would not be treated 102 (A) vs 22 (B) patients



patients may benefit from adjuvant chemotherapy [31]. One recent study showed an improvement in recurrence-free survival with the addition of immunotherapy to SBRT [6] and prospective studies are ongoing (NCT03924869, NCT03050554, NCT03833154). It would be of great interest to have a non-invasive tool to select patients who would benefit from this additional treatment to avoid immunotherapy-induced side effects [6].

Two studies have focused on clinical parameters and have developed several scores to predict distant recurrence which included age, histology, or tumor stage as in our study [7, 32]. Gao et al. developed a multicenter model based on clinical parameters and standard PET features (SUV_{max}) with internal and external validation to predict distant metastasis. The model identified a subgroup of high-risk patients who had increased rates of metastatic spread in the training set (23.9% [17 of 71] versus 12.5% [119 of 953], $p=0.006$), the internal test set (26.7% [4 of 15] versus 13.3% [32 of 241], $p=0.148$), and the external validation dataset (28.6% [4 of 14] versus 16.4% [19

of 116], $p=0.259$). This model has a moderate sensitivity and is not significant in the testing sets. In our study, the clinical model that yielded the best results includes 3 clinical parameters also present in their model. Clinical and radiomic models identified a high-risk patient subgroup with increased rates of distant metastases: (i) for the clinical model, in the training set (37.1% [59 of 159] versus 13.2% [15 of 114], $p<0.0001$), the internal test set (32.8% [43 of 131] versus 25.0% [15 of 60], $p=0.36$), and the external testing set (20.7% [6 of 29] versus 14.7% [5 of 34], $p=0.77$); (ii) and for the radiomic model, in the training set (48.5% [66 of 136] versus 5.8% [8 of 137], $p<0.0001$), the internal test set (54.4% [49 of 90] versus 8.9% [9 of 101], $p<0.0001$), and the external testing set (37.0% [10 of 27] versus 2.8% [1 of 36], $p=0.0014$). In comparison, the radiomic model outperforms the clinical model to predict the occurrence of distant metastases. Moreover, we have chosen as endpoint recurrence at any time and not at 1 year, since the median time to recurrence is around 20 months (table S2) [7].

Several studies have investigated the prognostic factors of patients who received SBRT for ES-NSCLC using radiomic analysis on [^{18}F]FDG PET/CT [33] or CT scans [34]. However, very few studies have focused on regional and/or distant recurrence [17, 35–37]. The [^{18}F]FDG PET/CT studies were retrospective and monocentric, and none included a training–testing scheme or an external validation cohort. In addition, most of these studies considered a relatively small number of radiomic features that did not conform to IBSI guidelines and were based on different segmentation methods or intensity discretization schemes, which may partly explain why different features were identified in these studies. Reproducibility and comparison between radiomic studies not conforming to IBSI guidelines are not feasible [28]. In contrast, Yu et al. retrospectively studied the predictive value of radiomic parameters extracted from contrast-enhanced CT images in 2 cohorts of stage I NSCLC patients corresponding to a training cohort ($n = 147$) and an independent validation cohort ($n = 295$) [38]. They developed a model to predict overall survival and the risk of distant metastatic recurrence. Their model was composed of two radiomic features. However, the 2 cohorts had not received the same treatment: patients in the training cohort had received surgery, while those in the validation cohort had received SBRT. In addition, the cohorts showed a significant difference in patient age, which may explain the difference in treatment. These potential biases may have an impact on the validity and reproducibility of the model. In another study, Sawayanagi et al. demonstrated the prognostic value of GTV-derived GLSZM features on the pretherapy CT image on OS in patients with localized NSCLC treated with curative SBRT [39]. In contrast to these studies, the present study was designed with a much larger sample size and, more importantly, with external testing sets treated by SBRT. Another important point is the reproducibility of our best model from analog PET/CT to digital PET/CT. Digital PET/CT has improved diagnostic accuracy, while reducing radiation dose and examination time, making it essential to validate radiomic models on this new technology of PET/CT, which will be equipping the majority of centers over the next few years [40].

We also developed models for distant recurrence and isolated regional recurrence, as we could discuss different management for this last group of patients [41]. The subgroup of patients with isolated regional recurrence seems to have a better prognosis, and we could avoid systemic treatment and propose close monitoring to offer them early localized salvage treatment (supplemental data E).

In our study, the three radiomic features retained in the radiomic model are extracted from PET imaging and are robust to inter-reader segmentation variations (intra-class correlation > 0.80). The most important, `original_first-order_Entropy` (implied in 48% in the radiomic model), is a

first-order feature which classifies heterogeneous tumors as the most likely to develop regional or/and distant recurrence. The other 2 are textural features (`original_gldm_LowGray-LevelEmphasis` and `original_gldm_DifferenceAverage`) one of which (`original_gldm_DifferenceAverage`) is a parameter already highlighted in a previous study (feature called dissimilarity in this study) carried out on part of the University Hospital of Liège cohort for the risk of recurrence and cancer-specific survival [14]. The combined model performed less well than the radiomic model alone, due to the poorer performance of the clinical parameter (tumor stage) incorporated into the model. No dosimetric parameter was predictive of the risk of regional and/or distant recurrence. This result is consistent with the literature, where dosimetric parameters are correlated only with local control but not with the risk of relapse outside radiotherapy fields [5, 42].

Our study has several limitations. Its retrospective nature can lead to several biases, including the lack of histologic confirmation of the primary tumor in few cases as is the case in many studies based on SBRT-treated lung cancer patients. Nevertheless, each case was evaluated in a multidisciplinary committee to minimize this bias. The 2 external testing sets had shorter follow-up than the Liège cohort, which may explain the lower recurrence rate and the lowest number of cancer-specific deaths. Longer-term evaluation of the radiomic model on these 2 cohorts would be interesting. Our model had a high sensitivity but a poor specificity; it means that the model is more prone to detect not only the disease even when this is not present but also a higher rate of false positives, resulting in a model still not able to identify true positives only. However, from a clinical point of view, it seems more interesting to have a better sensitivity than specificity to treat almost all patients at risk of recurrence. Neural networks are often described as “black boxes,” making the models generated difficult to understand [43]. However, in our study, our best model was composed of only 3 radiomic features whose respective contributions are available to make the model comprehensible and generate hypotheses about its expression of tumor heterogeneity. Another important issue in radiomic studies is the correlation of radiomic features with standard metrics (such as SUV measurements or metabolic volume) especially in PET. In our study, the 3 features composing our model have a low to moderate correlation with the standard parameters ($\rho = 0.29$ to 0.42). In radiomic studies, particularly those involving PET, volume, represented by the number of voxels, is probably the most important confounding factor. This problem concerns low-volume lesions, due to the limited spatial resolution of PET and the large-scale sampling of voxels, which means that texture analysis is performed on a small number of voxels, making the information provided by these texture features of limited relevance to volume alone. Several studies have addressed this issue, with the latest results showing that the

lower limit for exploiting texture features would be around 5 to 10 cm³, depending on the features and methodology adopted in the radiomic workflow, notably the gray-scale discretization method or the design of texture matrices [43, 44]. In our study, the range of considered volumes was 4–60 cm³ with rather large mean and median values of 11 and 8 cm³ respectively, which corresponds to 500 voxels for the smallest volume but a mean and median number of 5800 and 2500 voxels respectively. Despite these limitations, our study scores 53% (19 out of 36 items) on the radiomic quality score (Supplemental data F), which places it in favorable comparison with the greater part of prior radiomic studies.

Conclusion

Radiomic features extracted from [¹⁸F]FDG PET/CT outperformed clinical variables in selecting patients with ES-NSCLC treated with SBRT at high risk of regional or distant recurrence and could be useful to better individualize patients who would benefit from adjuvant systemic treatment. Prospective validation is necessary.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00259-023-06510-y>.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval All procedures were performed in accordance with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The study design and exemption from informed consent were approved by the Institutional Review Board of Liege University Hospital.

Consent to participate The study design and exemption from informed consent were approved by the Institutional Review Boards of Liege University Hospital (2022/285) and Brest University Hospital (29BRC16.0147).

Competing interests The authors declare no competing interests.

References


- Nagata Y, Hiraoka M, Shibata T, Onishi H, Kokubo M, Karasawa K, et al. Prospective trial of stereotactic body radiation therapy for both operable and inoperable T1N0M0 non-small cell lung cancer: Japan Clinical Oncology Group study JCOG0403. *Int J Radiat Oncol Biol Phys*. 2015;93:989–96.
- Guckenberger M, Allgauer M, Appold S, Dieckmann K, Ernst I, Ganswindt U, et al. Safety and efficacy of stereotactic body radiotherapy for stage I non-small-cell lung cancer in routine clinical practice: a patterns-of-care and outcome analysis. *J Thorac Oncol*. 2013;8:1050–8.
- Ricardi U, Frezza G, Filippi AR, Badellino S, Levis M, Navarria P, et al. Stereotactic ablative radiotherapy for stage I histologically proven non-small cell lung cancer: an Italian multicenter observational study. *Lung Cancer*. 2014;84:248–53.
- Timmerman R, Paulus R, Galvin J, Michalski J, Straube W, Bradley J, et al. Stereotactic body radiation therapy for inoperable early stage lung cancer. *JAMA*. 2010;303:1070–6.
- Eriguchi T, Takeda A, Nemoto T, Tsurugai Y, Sanuki N, Tateishi Y, et al. Relationship between dose prescription methods and local control rate in stereotactic body radiotherapy for early stage non-small-cell lung cancer: systematic review and meta-analysis. *Cancers (Basel)*. 2022;14:3815.
- Chang JY, Lin SH, Dong W, Liao Z, Gandhi SJ, Gay CM, et al. Stereotactic ablative radiotherapy with or without immunotherapy for early-stage or isolated lung parenchymal recurrent node-negative non-small-cell lung cancer: an open-label, randomised, phase 2 trial. *Lancet*. 2023;S0140–6736(23):01384–93.
- Gao SJ, Jin L, Meadows HW, Shafman TD, Gross CP, Yu JB, et al. Prediction of distant metastases after stereotactic body radiation therapy for early stage NSCLC: development and external validation of a multi-institutional model. *J Thorac Oncol*. 2023;18:339–49.
- Vaz SC, Adam JA, Delgado Bolton RC, Vera P, van Elmpt W, Herrmann K, et al. Joint EANM/SNMMI/ESTRO practice recommendations for the use of 2-[¹⁸F]FDG PET/CT external beam radiation treatment planning in lung cancer V1.0. *Eur J Nucl Med Mol Imaging*. 2022;49:1386–406.
- Sharma A, Mohan A, Bhalla AS, Sharma MC, Vishnubhatla S, Das CJ, et al. Role of various metabolic parameters derived from baseline 18F-FDG PET/CT as prognostic markers in non-small cell lung cancer patients undergoing platinum-based chemotherapy. *Clin Nucl Med*. 2018;43:e8–17.
- Kwon W, Howard BA, Herndon JE, Patz EF. FDG uptake on positron emission tomography correlates with survival and time to recurrence in patients with stage I non-small-cell lung cancer. *J Thorac Oncol*. 2015;10:897–902.
- Na F, Wang J, Li C, Deng L, Xue J, Lu Y. Primary tumor standardized uptake value measured on F18-fluorodeoxyglucose positron emission tomography is of prediction value for survival and local control in non-small-cell lung cancer receiving radiotherapy: meta-analysis. *J Thorac Oncol*. 2014;9:834–42.
- Hoang JK, Hoagland LF, Coleman RE, Coan AD, Herndon JE, Patz EF. Prognostic value of fluorine-18 fluorodeoxyglucose positron emission tomography imaging in patients with advanced-stage non-small-cell lung carcinoma. *J Clin Oncol*. 2008;26:1459–64.
- Agarwal M, Brahmanday G, Bajaj SK, Ravikrishnan KP, Wong C-YO. Revisiting the prognostic value of preoperative (18)F-fluoro-2-deoxyglucose ((18)F-FDG) positron emission tomography (PET) in early-stage (I & II) non-small cell lung cancers (NSCLC). *Eur J Nucl Med Mol Imaging*. 2010;37:691–8.
- Lovinfosse P, January ZL, Coucke P, Jodogne S, Bernard C, Hatt M, et al. FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy. *Eur J Nucl Med Mol Imaging*. 2016;43:1453–60.
- Dissaux G, Visvikis D, Da-Ano R, Pradier O, Chajon E, Barillot I, et al. Pretreatment 18F-FDG PET/CT radiomics predict local recurrence in patients treated with stereotactic body radiotherapy for early-stage non-small cell lung cancer: a multicentric study. *J Nucl Med*. 2020;61:814–20.
- Wu J, Aguilera T, Shultz D, Gudur M, Rubin DL, Loo BW, et al. Early-stage non-small cell lung cancer: quantitative imaging characteristics of (18)F fluorodeoxyglucose PET/CT allow prediction of distant metastasis. *Radiology*. 2016;281:270–8.
- Oikonomou A, Khalvati F, Tyrrell PN, Haider MA, Tarique U, Jimenez-Juan L, et al. Radiomics analysis at PET/CT contributes

- to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy. *Sci Rep.* 2018;8:4003.
18. Pyka T, Bundschuh RA, Andratschke N, Mayer B, Specht HM, Papp L, et al. Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy. *Radiat Oncol.* 2015;10:100.
 19. Opitz D, Maclin R. Popular ensemble methods: an empirical study. *JAIR.* 1999;11:169–98.
 20. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. *Radiology.* 2018;286:887–96.
 21. Bourbonne V, Lucia F, Jaouen V, Bert J, Rehn M, Pradier O, et al. 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;164:43–9.
 22. Janvary ZL, Jansen N, Baart V, Devillers M, Dechambre D, Lenaerts E, et al. Clinical outcomes of 130 patients with primary and secondary lung tumors treated with Cyberknife robotic stereotactic body radiotherapy. *Radiol Oncol.* 2017;51:178–86.
 23. Senthil S, Lagerwaard FJ, Haasbeek CJA, Slotman BJ, Senan S. Patterns of disease recurrence after stereotactic ablative radiotherapy for early stage non-small-cell lung cancer: a retrospective analysis. *Lancet Oncol.* 2012;13:802–9.
 24. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials.* 1996;17:343–6.
 25. Belli ML, Mori M, Broggi S, Cattaneo GM, Bettinardi V, Dell’Oca I, et al. Quantifying the robustness of [18F]FDG-PET/CT radiomic features with respect to tumor delineation in head and neck and pancreatic cancer patients. *Phys Med.* 2018;49:105–11.
 26. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep.* 2013;3:3529.
 27. Radiomic features — PyRadiomics v3.1.0rc2.post5+g6a761c4 documentation. <https://pyradiomics.readthedocs.io/en/latest/feature.html>. Accessed 29 Jun 2023.
 28. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology.* 2020;295:328–38.
 29. Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage.* 2018;167:104–20.
 30. Caetano SJ, Sonpavde G, Pond GR. C-statistic: a brief explanation of its construction, interpretation and limitations. *Eur J Cancer.* 2018;90:130–2.
 31. Ernani V, Appiah AK, Marr A, Zhang C, Zhen W, Smith LM, et al. Adjuvant systemic therapy in patients with early-stage NSCLC treated with stereotactic body radiation therapy. *J Thorac Oncol.* 2019;14:475–81.
 32. Zhou Z, Folkert M, Cannon N, Iyengar P, Westover K, Zhang Y, et al. Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters. *Radiother Oncol.* 2016;119:501–4.
 33. Manafi-Farid R, Askari E, Shiri I, Pirich C, Asadi M, Khateri M, et al. [18F]FDG-PET/CT radiomics and artificial intelligence in lung cancer: technical aspects and potential clinical applications. *Semin Nucl Med.* 2022;52:759–80.
 34. Fornaçon-Wood I, Faivre-Finn C, O’Connor JPB, Price GJ. Radiomics as a personalized medicine tool in lung cancer: Separating the hope from the hype. *Lung Cancer.* 2020;146:197–208.
 35. Hao H, Zhou Z, Wang J. Distant failure prediction for early stage NSCLC by analyzing PET with sparse representation. In: *Medical Imaging 2017: Computer-Aided Diagnosis.* SPIE; 2017. p. 1008–14.
 36. Li H, Galperin-Aizenberg M, Pryma D, Simone CB, Fan Y. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol.* 2018;129:218–26.
 37. Li S, Yang N, Li B, Zhou Z, Hao H, Folkert MR, et al. A pilot study using kernelled support tensor machine for distant failure prediction in lung SBRT. *Med Image Anal.* 2018;50:106–16.
 38. Yu W, Tang C, Hobbs BP, Li X, Koay EJ, Wistuba II, et al. Development and validation of a predictive radiomics model for clinical outcomes in stage I non-small cell lung cancer. *Int J Radiat Oncol Biol Phys.* 2018;102:1090–7.
 39. Sawayanagi S, Yamashita H, Nozawa Y, Takenaka R, Miki Y, Morishima K, et al. Establishment of a prediction model for overall survival after stereotactic body radiation therapy for primary non-small cell lung cancer using radiomics analysis. *Cancers (Basel).* 2022;14:3859.
 40. Lucia F, Bourbonne V, Pleyers C, Dupré P-F, Miranda O, Visvikis D, et al. Multicentric development and evaluation of 18F-FDG PET/CT and MRI radiomics models to predict para-aortic lymph node involvement in locally advanced cervical cancer. *Eur J Nucl Med Mol Imaging.* 2023. <https://doi.org/10.1007/s00259-023-06180-w>.
 41. Lee TH, Shin H, Ahn YC, Kang MK, Song C, Kim WC, et al. Regional lymph node recurrence after stereotactic body radiation therapy for lung cancer: patterns of recurrence, treatment approaches, and clinical outcomes (KROG 21–09). *Radiother Oncol.* 2023;183: 109572.
 42. Klement RJ, Sonke J-J, Allgäuer M, Andratschke N, Appold S, Belderbos J, et al. Correlating dose variables with local tumor control in stereotactic body radiation therapy for early-stage non-small cell lung cancer: a modeling study on 1500 individual treatments. *Int J Radiat Oncol Biol Phys.* 2020;107:579–86.
 43. Hatt M, Cheze Le Rest C, Antonorsi N, Tixier F, Tankyevych O, Jaouen V, et al. Radiomics in PET/CT: current status and future AI-based evolutions. *Semin Nucl Med.* 2021;51:126–33.
 44. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging.* 2017;44:151–65.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

François Lucia^{1,2,3,4}  · Thomas Louis³ · François Cousin³ · Vincent Bourbonne^{1,2} · Dimitris Visvikis² · Carole Mievis⁵ · Nicolas Jansen⁵ · Bernard Duysinx⁶ · Romain Le Pennec^{7,8} · Malik Nebbache¹ · Martin Rehn¹ · Mohamed Hamya¹ · Margaux Geier⁹ · Pierre-Yves Salaun^{7,8} · Ulrike Schick^{1,2} · Mathieu Hatt² · Philippe Coucke⁵ · Roland Hustinx^{3,10} · Pierre Lovinfosse³

✉ François Lucia
francois.lucia@chu-brest.fr

¹ Radiation Oncology Department, University Hospital, Brest, France

² LaTIM, INSERM, UMR 1101, Univ Brest, Brest, France

³ Division of Nuclear Medicine and Oncological Imaging, University Hospital of Liège, Liège, Belgium

⁴ Service de Radiothérapie, CHRU Morvan, 2 Avenue Foch, 29609 Cedex Brest, France

⁵ Department of Radiotherapy Oncology, University Hospital of Liège, Liège, Belgium

⁶ Division of Pulmonology, CHU Liège, Liège, Belgium

⁷ Nuclear Medicine Department, University Hospital, Brest, France

⁸ GETBO, INSERM, UMR 1304, University of Brest, UBO, Brest, France

⁹ Medical Oncology Department, University Hospital, Brest, France

¹⁰ GIGA-CRC In Vivo Imaging, University of Liège, Liège, Belgium