Supplemental Table 1: Dosimetric data

		Training set N = 64	Testing set N = 23
Median dose (range)(Gy)		54 (30 - 60)	48 (48 - 48)
Median BED (range)(Gy)		150 (45 - 180)	105.6 (105.6 - 105.6)
Median dose per fraction (range)(Gy)		15 (5 - 20)	4 (4 - 4)
Median fraction number (range)(Gy)		4 (3 - 8)	12 (12 - 12)
Median volume (range)(cm^3)		133.3 (8.3 - 946.9)	149.7 (5.6 - 987.2)
Treatment units			
Cyberknife		38	0
Truebeam		18	0
Truebeam Stx Novalis		8	23
Number of Fractions	Dose/fraction (Gy)		
3	< 18	1	0
	=18	5	0
	=20	23	0
	>18 (except 20)	1	0
4	< 12	0	0
	=12	17	23
	>12	5	0
5	<10	0	0
	=10	2	0
	>10	1	0
6	< 9	2	0
	=9	6	0
	>9	0	0
8	=7,5	1	0

Abbreviations: BED=biological equivalent dose

Supplemental Table 2: Report on image processing and image biomarker extraction

General					
Imaging	PET, CT				
Acquisition and reconstruction	**Acquisition parameters**	**Biograph – Siemens (Brest – Nantes)**		**Discovery 690 – General Electric (Tours)**	
		PET	CT	PET	CT
	^{18}F-FDG activity (MBq)*	350–550	–	350–550	–
	Min/bed position	2.5	–	2	–
	Crystal	LSO	–	LYSO	–
	Reconstruction	Iterative	–	Iterative, TOF Sharp IR	–
	Matrix (pixels)	200x200	512×512	256×256	512×512
	Resolution (mm)	4.07x4.07	0.98×0.98	2.73×2.73	0.98x0.98
	Slice thickness (mm)	2.0	2.0	3.27	3.27
	Slices	–		–	
	Voltage (kV)	–	100	–	140
	Tube current (mA)	–	95	–	140
	Reconstruction Method	PSF, TOF2i21s	–	VPFXS	–
	Correction Applied	Norm,dtim,attn scat,decy,ran		Decy,attn,scat, dtim,ransng, dcal,slsens,norm	
	*Administered activity was calculated according to the European Association of Nuclear Medicine (EANM) guidelines 1.0 and, from February 2015, 2.0 [21]				
	Acquisition parameters	**Ingenuity –Philips (Tours)**		**Discovery ST – General Electric (Rennes - Nantes)**	
		PET	CT	PET	CT
	^{18}F-FDG activity (MBq)*	350–550	–	350–550	–
	Min/bed position	2.5	–	2	–
	Crystal	LSO	–	LYSO	–
	Reconstruction	Iterative	–	Iterative, TOF Sharp IR	–
	Matrix (pixels)	144x144	512×512	128x128	512×512
	Resolution (mm)	4x4	0.98x0.98	5.5x5.5	0.98x0.98
	Slice thickness (mm)	4	0.5	3.27	2.5
	Slices	–		–	
	Voltage (kV)	–	140	–	140
	Tube current (mA)	–	58	–	120
	Reconstruction	BLOB-OS-TF	–	3D IR	–
	Correction Applied	Decy,radl,attn, scat,dtim, ran,norm,cln		Decy,attn,scat, dtim,ran,dcal, slsens,norm	
Approach	The images were analysed as a volume (3D).				
Process structure	Image acquisition -> reconstruction -> anonymisation -> segmentation -> export -> radiomics analysis -> feature calculation report				

Software	MIRAS software V 1.06 (LaTIM INSERM, UMR 1101, Brest, France)
Data availability	The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request for research purposes.
Data conversion	
Procedure	None
Image post-acquisition processing	
Procedure	None
Segmentation	
ROI	The volume of interest (VOI) included the primary tumour lesion.
Procedure	The ROIs were semi-automatically defined on PET images with the fuzzy locally adaptive Bayesian (FLAB) software, and manually on CT images with MiM Maestro® software. (MiM software Inc Cleveland, OH 44122).
Interpolation	
Voxel dimensions	None. Original dimensions were kept for all images.
Image interpolation method	Not applicable
Image intensity rounding	Not applicable
ROI interpolation method	Not applicable
ROI partial volume	Not applicable
Re-segmentation	
ROI mask criteria	None
Discretisation	
Discretisation method	PET: fixed number of bins, 64 bins CT: fixed number of bins, 64 bins

Feature calculation	
Feature set	**I.INTENSITY HISTOGRAMM FEATURES** Minimum Maximum Mean Variance Standard deviation Skewness Kurtosis Energy Entropy **II.THREE-DIMENSIONAL SHAPE** Volume Approximate volume 3D surface Ration 3D volume Compactness V1 Compactness V2 Spherical disproportion Sphericity Asphericity Maximun 3D diameter Major axis length Minor axis length Least axis length Elongation Flatness **III.SECOND ORDER STATISTICS FEATURES DERIVED FROM CO-OCCURRENCE MATRIX AND** **DIFFERENCE GREY LEVEL MATRIX** A. **Co-occurrence matrix (GLCM (Grey Level Co-occurrence Matrix))** Max co-occurrence Average co-occurrence Variance co-occurrence Entropy co-occurrence Difference Average Difference Variance Difference Entropy

	Sum Average Sum Variance Sum Entropy Angular second moment Contrast Dissimilarity Inverse difference Inverse difference Normalized Inverse difference moment Inverse variance Correlation Autocorrelation Tendency Shade Prominence First Measure of Information Correlation Second Measure of Information Correlation **B. Difference grey level matrix** Coarseness Contrast Busyness Complexity Strength **IV. TEXTURAL FEATURES DERIVED FROM ZONE SIZE AND ALIGMENT MATRIX** **A. Alignment matrix** Short Run Emphasis Long Run Emphasis Grey-level non-uniformity Run length non-uniformity Run percentage Low Grey Level Run Emphasis High Grey Level Run Emphasis Grey-level non-uniformity normalized Run length non-uniformity normalized Grey-level Variance Run-Length Variance **B. Zone size matrix** Small zone emphasis Large Zone Emphasis Low grey level zone emphasis

	High grey level zone emphasis Small zone low grey level emphasis Small zone high grey level emphasis Large zone low grey level emphasis Large zone high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalized Zone size non-uniformity Zone size entropy Zone size non-uniformity normalized Grey level variance Zone size variance
Feature parameters	Texture matrices are built in 3D following the merging strategy (see IBSI reference document).
Standardisation	Features values have been checked with the most up-to-date consensus of the IBSI benchmark values.

Supplemental Table 3: PET and CT features with an AUC > 0.7 in the training set

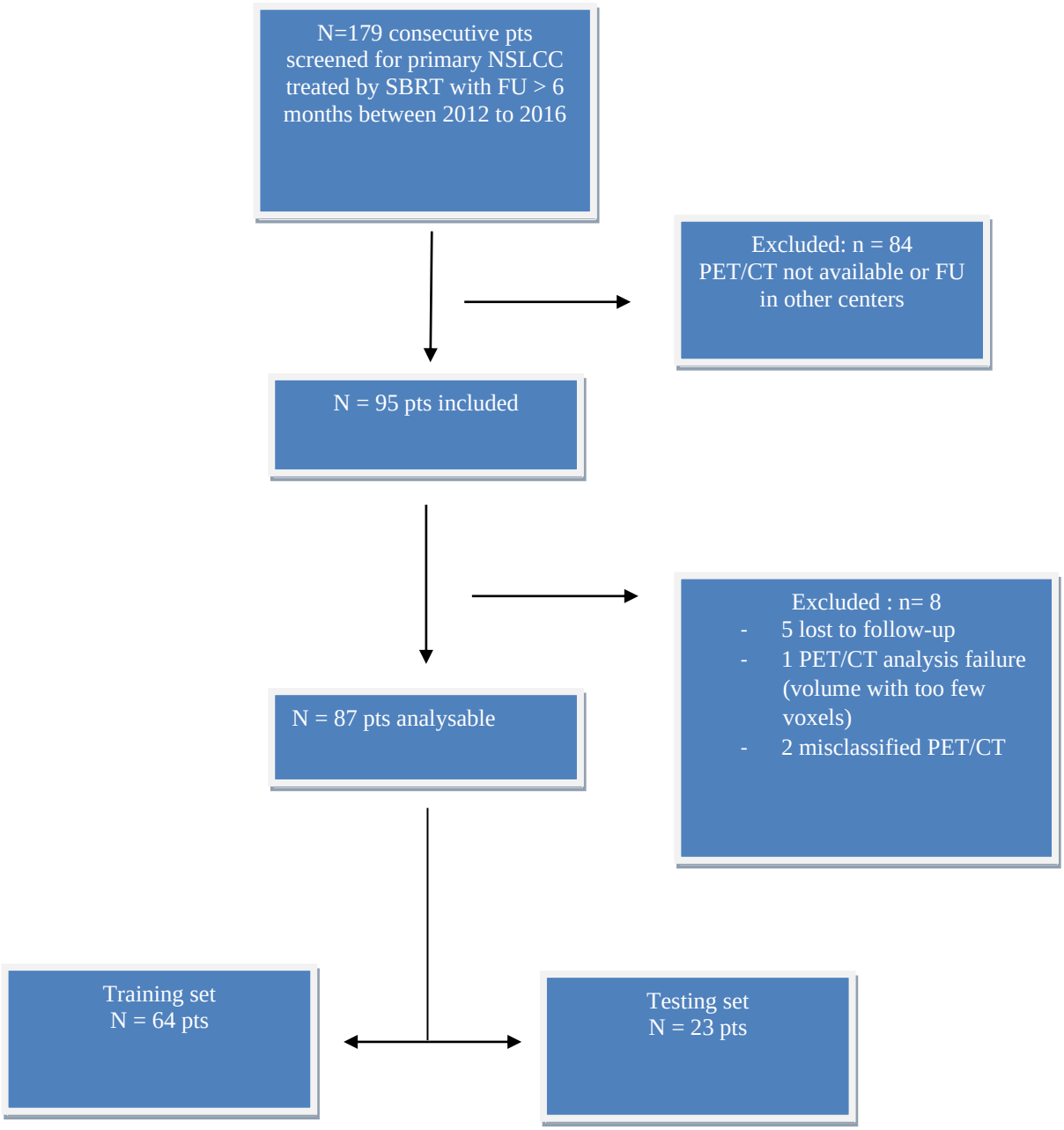
Variable (AUC > 0.7)	Hazard ratio	95% CI	*p* value
Radiomics PET :			
IC2 (AUC 0.83, Se 1.0 - Sp 0.72)	undefined	undefined	**0.005**
Strength (AUC 0.86, Se 1.0 - Sp 0.72)	undefined	undefined	**0.001**
Radiomics CT:			
Flatness (AUC 0.93, Se 1.0 - Sp 0.88)	undefined	undefined	**<0.001**
Shade (AUC 0.81, Se 0.75 - Sp 0.88)	13.4	(1.1-168)	**0.003**
Elongation (AUC 0.79, Se 1.0 - Sp 0.69)	undefined	undefined	0.022

Abbreviations : Se=sensitivity, Sp=specificity, AUC= area under the curve, CI=confidence interval

Supplemental Table 4: Spearman's rank correlation between variables

Variables	IC2 PET	Strength PET	Flatness CT	Shade CT	Elongation CT
Age (year)	-0.003	0.003	0.004	0.344	-0.01
Tumor volume (cm^3)	-0.585	-0.636	-0.009	-0.032	0.215
SUV_{max}	-0.023	-0.093	-0.039	0.118	-0.105
IC2 PET	1	0.728	-	-	-
Strength PET	-	1	-	-	-
Flatness CT	-	-	1	0.296	0.636
Shade CT	-	-	0.296	1	0.197
Elongation CT	-	-	0.636	0.197	1

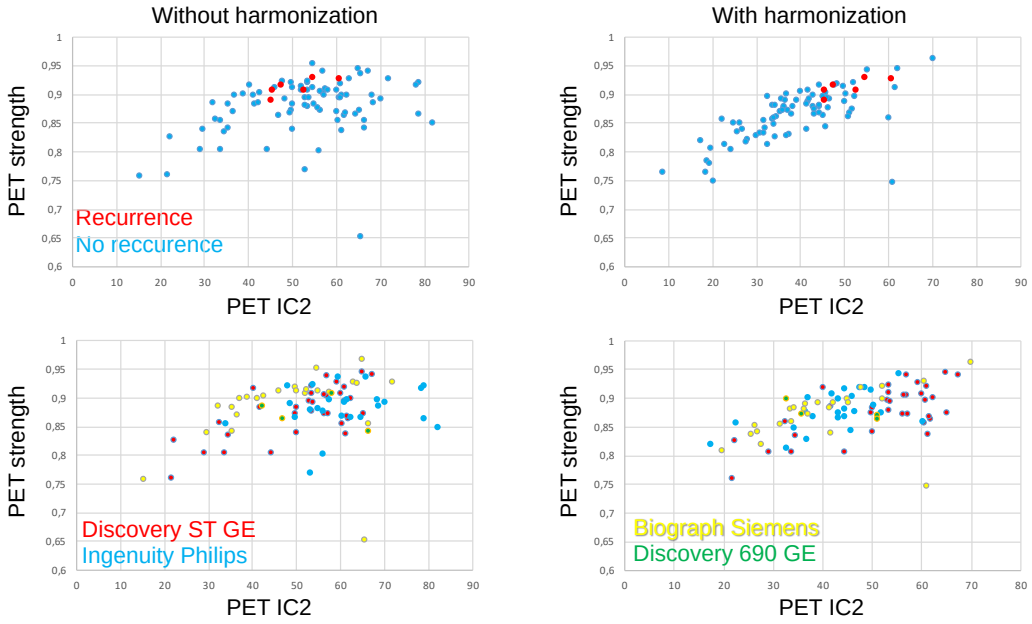
Supplemental Figure 1: Flow chart of patients selection



Supplemental Table 5: Patient's center data

	Brest	Rennes	Tours	Nantes
Inclusion's duration	8 months	52 months	50 months	46 months
Start – End inclusion	Apr 2016 – Dec 2016	Apr 2012 – Aug 2016	June 2012 – Aug 2016	Jan 2012 – Nov 2015
Median Follow-up (months)	15.7 (4 – 23)	27 (5 – 58)	20 (2 – 63)	25 (7 – 58)

Supplemental Figure 2: IC2 $_{PET}$ and Strength $_{PET}$ features scatter plot before and after ComBat harmonization according to the different PET/CT systems.



Supplemental Table 6: Accuracy (sensitivity and specificity) of PET and CT features with and without harmonization in training and testing set

Variable		Without harmonization	
		Training set	Testing set
PET			
	IC2	0.78 (Se 0.75 - Sp 0.93)	0.68 (Se 0.5 - Sp 0.86)
	Strength	0.64 (Se 1.0 - Sp 0.57)	0.88 (Se 1.0 - Sp 0.76)
	IC2 + Strength	**0.85 (Se 0.75 - Sp 0.95)**	**0.95 (Se 1.0 - Sp 0.90)**
CT			
	Flatness	0.93 (Se 0.75 - Sp 1.0)	0.50 (Se 0.0 - Sp 1.0)
	Shade	0.77 (Se 1.0 - Sp 0.48)	0.80 (Se 1.0 - Sp 0.6)
	Elongation	0.724 (Se 1.0 - Sp 0.60)	0.33 (Se 0.0 - Sp 1.0)
PET/CT			
	IC2 + Flatness	0.75 (Se 0.5 - Sp 0.85)	0.5 (Se 0.0 - Sp 1.0)

Variable		With harmonization	
		Training set	Testing set
PET			
	IC2	0.83 (Se 1.0 - Sp 0.72)	0.83 (Se 1.0 - Sp 0.67)
	Strength	0.86 (Se 1.0 - Sp 0.72)	0.88 (Se 1.0 - Sp 0.76)
	IC2 + Strength	**0.94 (Se 1.0 - Sp 0.88)**	**0.91 (Se 1.0 - Sp 0.81)**
CT			
	Flatness	0.93 (Se 1.0 - Sp 0.88)	0.40 (Se 0.0 - Sp 0.8)
	Shade	0.81 (Se 0.75 - Sp 0.88)	0.40 (Se 0.0 - Sp 1.0)
	Elongation	0.79 (Se 1.0 - Sp 0.69)	0.275 (Se 0.0 - Sp 1.0)
PET/CT			
	IC2 + Flatness	0.98 (Se 1.0 - Sp 0.96)	0.45 (Se 0.0 - Sp 1.0)

Supplemental Table 7: Accuracy results for the model combining the two PET features (with cut-off values of 0.89 and 45.11 for IC2 and Strength respectively)

	Training set	Testing set
With histology	0.97 (Se 1 and Sp 0.97)	0.94 (Se 1 and Sp 0.94)
Without histology	0.83 (Se 1 and Sp 0.81)	1.0 (Se 1 and Sp 1)

Additional details about the ComBat methodology used for harmonization

The Combat harmonization was initially proposed for correcting the so-called "batch effect" in genomic studies (*2*). ComBat determines an appropriate transformation for each feature through Bayes estimates in the entire feature space, based on the *a priori* "batch" effect observed on feature values. When necessary, the features values are modified so their distribution better match. As a result, most feature values are modified to an arbitrary new reference. This batch label is user defined and in our case it was set as each combination of PET/CT scanner model, acquisition protocol and reconstruction settings (see supplemental table 2). ComBat has been shown to outperform other similar harmonization statistical methods and to be robust for small samples (*3*). We applied ComBat without accounting for any biological covariate as there was no difference between cohorts in terms of clinical or histopathological parameters. It should be emphasized that ComBat is applied as a pre-processing step to the entire dataset (all features from all 4 centers) before any statistical analysis (correlation, training the models, and testing evaluation) is carried out. ComBat was applied only to radiomic features, not on other clinical variables.

In order to further evaluate (*e.g.*, in a prospective study or in another external testing set) our radiomic model trained and validated on multicentric dataset harmonized with ComBat, the features from the patients of the new cohort should be added to the existing database and harmonized with ComBat. Finally, to use the model clinically on new patients, a similar process can be followed. The new patient radiomic features are added to the database so that the features are transformed and the previously built model can be applied to obtain the prediction for that patient. If the scanner, the acquisition protocol and/or the reconstruction settings are modified for a specific center, then a small sample of patients will have to be collected so to re-harmonize the features with those used for the training/validation of the model.

REFERENCES:

1. Alex Zwanenburg SL, Martin Vallières, Steffen Löck. Image biomarker standardisation initiative - feature definitions. 2017.

2. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8:118-127.

3. Chen C, Grennan K, Badner J, et al. Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One.* 2011;6:e17238.

4. Vallieres M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible Radiomics Research for Faster Clinical Translation. *J Nucl Med.* 2018;59:189-193.