Article

Value of Shape and Texture Features from ^{18}F-FDG PET/CT to Discriminate between Benign and Malignant Solitary Pulmonary Nodules: An Experimental Evaluation

Barbara Palumbo [1,†] , Francesco Bianconi [2,*,†] , Isabella Palumbo [3] , Mario Luca Fravolini [2] , Matteo Minestrini [1] , Susanna Nuvoli [4] , Maria Lina Stazza [4] , Maria Rondini [4] and Angela Spanu [4]

1 Section of Nuclear Medicine and Health Physics, Department of Surgical and Biomedical Sciences, Università degli Studi di Perugia, Piazza Lucio Severi 1, 06132 Perugia, Italy; barbara.palumbo@unipg.it (B.P.); matteo.minestrini@ospedale.perugia.it (M.M.)
2 Department of Engineering, Università degli Studi di Perugia, Via Goffredo Duranti 93, 06135 Perugia, Italy; mario.fravolini@unipg.it
3 Section of Radiation Oncology, Department of Surgical and Biomedical Sciences, Università degli Studi di Perugia, Piazza Lucio Severi 1, 06132 Perugia, Italy; isabella.palumbo@unipg.it
4 Unit of Nuclear Medicine, Department of Medical, Surgical and Experimental Sciences, Università degli Studi di Sassari, Viale San Pietro 8, 07100 Sassari, Italy; snuvoli@uniss.it (S.N.); marialinastazza@yahoo.it (M.L.S.); maria.rondini01@ateneopv.it (M.R.); aspanu@uniss.it (A.S.)
* Correspondence: bianco@ieee.org; Tel.: +39-075-585-3706
† These authors contributed equally to this work.

Abstract: In this paper, we investigate the role of shape and texture features from ^{18}F-FDG PET/CT to discriminate between benign and malignant solitary pulmonary nodules. To this end, we retrospectively evaluated cross-sectional data from 111 patients (64 males, 47 females, age = 67.5 ± 11.0) all with histologically confirmed benign ($n = 39$) or malignant ($n = 72$) solitary pulmonary nodules. Eighteen three-dimensional imaging features, including conventional, texture, and shape features from PET and CT were tested for significant differences (Wilcoxon-Mann-Withney) between the benign and malignant groups. Prediction models based on different feature sets and three classification strategies (Classification Tree, k-Nearest Neighbours, and Naïve Bayes) were also evaluated to assess the potential benefit of shape and texture features compared with conventional imaging features alone. Eight features from CT and 15 from PET were significantly different between the benign and malignant groups. Adding shape and texture features increased the performance of both the CT-based and PET-based prediction models with overall accuracy gain being 3.4–11.2 pp and 2.2–10.2 pp, respectively. In conclusion, we found that shape and texture features from ^{18}F-FDG PET/CT can lead to a better discrimination between benign and malignant lung nodules by increasing the accuracy of the prediction models by an appreciable margin.

Keywords: solitary pulmonary nodule; shape; texture; radiomics

1. Introduction

A solitary pulmonary nodule (SPN) is usually defined as an "approximately round lesion that is less than 3 cm in diameter and that is completely surrounded by pulmonary parenchyma, without other abnormalities" [1]. With the widespread availability of chest CT and the improving capabilities of the imaging devices, the incidence of SPN has been rising constantly in recent years [2,3]. In the United States, the estimated detection rate (defined as at least one positive CT scan) increased from 3.9 per 1000 person-years in 2006 to 6.4 in 2012, of which the fraction of those who received a diagnosis of lung cancer within two years from the transcript was 5.2% [4]. Other works have reported estimates about the prevalence and malignancy rate of SPN between 2% and 69% [2,5,6] and between 7% and 40% [3,7,8], respectively.

Differential diagnosis of SPN includes benign aetiologies, such as pneumonia, fungi infection, tuberculosis, and hamartoma, as well as malignant ones—most commonly primary lung cancer, distant (metastatic) lesions, or lymphoma [7,9]. Assessment of the malignancy risk involves the evaluation of clinical and radiographic variables. Age, history of smoking, and exposure to carcinogenic agents are well-known risk factors, although their absence does not preclude malignancy [9]. Radiographic findings that are suggestive of benignity or malignancy include size, density, stability over time, margin appearance, wall thickness, and the presence of cavitation and calcification [9–11]. Clinical management of patients with SPN is determined on the basis of the risk assessed, and may involve routine CT follow-ups, functional imaging, and/or tissue sampling [7,9,12].

In recent years, quantitative extraction of imaging features from medical scans ("radiomics" [13–15]) has attracted widespread interest as a possible means to discriminate between benign vs. malignant SPN [16]. The rationale behind radiomics is to leverage on that fraction of image information which may have clinical relevance but go unnoticed to the human eye [17]. Radiomics also enables full-field analysis of the region of interest, while biopsies only capture a small portion of the lesion [18]. Several studies have proposed predictive models based on a range of combinations of radiomics features from CT, with overall reported accuracy between 70% and 95% [19–26]. Uptake parameters from ^{18}F fluodeoxyglucose (FDG) Positron Emission Tomography (PET henceforth) have also shown good diagnostic performance (accuracy between 65% and 91% in [3,8,27–30]) with potential improvements coming from the characterisation of uptake heterogeneity [31,32]. In a recent meta-analysis, Jia et al. [33] concluded that CT and PET/CT have both moderate-to-high diagnostic value in patients with SPN, with no significant differences between the two modalities.

Few studies, however, have addressed the problem of quantifying the gain that shape and texture features from CT and/or PET/CT may provide compared with conventional imaging features alone. Among them, Wu et al. [25] reported that adding texture features from CT to clinical and semantic variables could lead to a 0.03 increase of the area under the curve (AUC), whereas Balagurunathan et al. [23] determined that a combination of size, shape, and texture features could increment the AUC from 0.87 to 0.90 compared with a model based on longest diameter and volume of the nodule only.

In this work, we investigated whether shape and texture features from PET/CT could lead to a better discrimination between benign and malignant SPN compared with standard imaging features alone—that is, lesion size, density, and radiotracer uptake. To this end, we retrospectively evaluated the baseline PET/CT scans of 111 patients—all with histologically confirmed benign or malignant lesions—who had a positive transcript for SPN. The statistical analysis was carried out in two steps: we first screened the features for significance against benignity vs. malignancy, then used the significant features to build prediction models of different complexity.
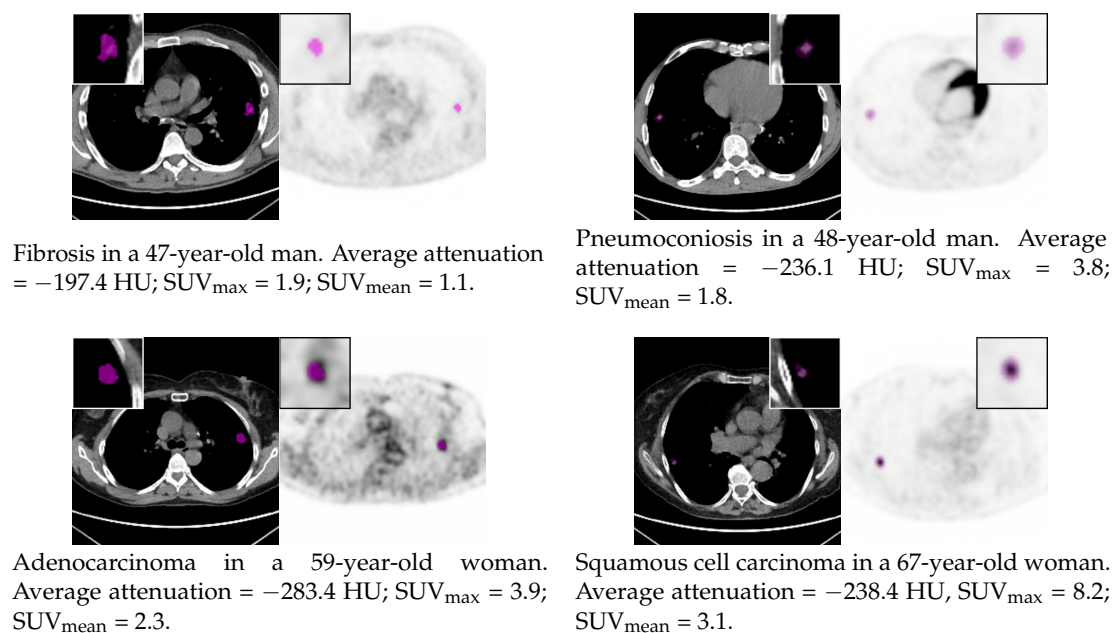
2. Materials and Methods

2.1. Study Population

We retrospectively evaluated a cohort of 111 patients (details of the study population in Table 1; commented sample scans in Figure 1) who underwent PET/CT examination for suspicious lung nodules at the Unit of Nuclear Medicine of the Università degli Studi di Sassari, Sassari, Italy, between November 2014 and May 2019. The inclusion criteria were: (1) presence of a clearly identifiable solid nodule at CT above 5 mm and up to 40 mm in maximum axial diameter (same criterion adopted in [29]), (2) no previous surgery, chemotherapy, and/or radiotherapy for the inspected lesion, and (3) histologically confirmed malignancy or benignity. The maximum axial diameter (volume) of the benign lesions at CT ranged between 7.4 mm and 35.8 mm (165.0 mm^3 and 11,222.2 mm^3), and that of the malignant ones between 10.0 mm and 38.1 mm (421.7 mm^3 and 18440.9 mm^3). Known risk factors in the patient series were: history of tuberculosis (prevalence in the study population = 5.4%), fibrosis (8.1%), pulmonary emphysema (8.1%), chronic obstructive pulmonary disease (17.1%), rheumatoid arthritis (5.4%), previous malignancy (14.4%), current smoking (23.4%), former smoking (25.2%), exposure to asbestos (5.4%), and exposure to other chemicals and/or agents (7.2%). The standard of reference for all the lesions was histological evaluation after bronchoscopy and/or surgical resection.

All patients gave written informed consent as part of the examination routine, and their data were treated according to the local privacy rules and regulations. Request for an ethical standard was waived due to the retrospective nature of the study.

Table 1. Characteristics of the patient series.

Attribute [Data Format]	Value
Demographics	
Age [mean ± SD]	67.5 ± 11.0
Female [*N* (%)]	47 (42.3)
Male [*N* (%)]	64 (57.7)
Histology	
Benign [*N* (%)]	39 (35.1)
Anthracosis [*N* (%)]	1 (0.9)
Fibrosis [*N* (%)]	10 (9.0)
Inflammation [*N* (%)]	26 (23.4)
Metaplasia [*N* (%)]	1 (0.9)
Pneumoconiosis [*N* (%)]	1 (0.9)
Malignant [*N* (%)]	72 (64.9)
Adenocarcinoma [*N* (%)]	46 (41.4)
Atypical carcinoid (NSCLC) [*N* (%)]	1 (0.9)
Metastasis [*N* (%)]	1 (0.9)
Neuroendocrine tumour [*N* (%)]	1 (0.9)
Small-cell lung cancer [*N* (%)]	2 (1.8)
Spinocellular carcinoma [*N* (%)]	4 (3.6)
Squamous cell carcinoma [*N* (%)]	9 (8.1)
Unspecified [*N* (%)]	8 (7.2)

Fibrosis in a 47-year-old man. Average attenuation = -197.4 HU; $SUV_{max} = 1.9$; $SUV_{mean} = 1.1$.

Pneumoconiosis in a 48-year-old man. Average attenuation = -236.1 HU; $SUV_{max} = 3.8$; $SUV_{mean} = 1.8$.

Adenocarcinoma in a 59-year-old woman. Average attenuation = -283.4 HU; $SUV_{max} = 3.9$; $SUV_{mean} = 2.3$.

Squamous cell carcinoma in a 67-year-old woman. Average attenuation = -238.4 HU, $SUV_{max} = 8.2$; $SUV_{mean} = 3.1$.

Figure 1. Sample CT and PET scans showing benign (first row) and malignant (second row) pulmonary nodules. Magnified (2×) views of the manually segmented lesions are shown in the insets; basic CT and PET parameters are also reported.

2.2. Acquisition Protocol

All the patients fasted for at least 6 h and were checked for blood glucose level <150 mg/dL before examination. Afterwards, they were administered an intravenous bolus of 3.75 Mbq/Kg, and whole-body image acquisition started 60 min thereafter. PET images of size 256 px × 256 px (slice thickness 3.27 mm, in-plane pixel spacing ≈2.73 mm in both directions) were acquired in helicoidal mode and reconstructed via an iterative reconstruction algorithm (GE-VPFXS). Computed tomography scans for attenuation correction were also acquired in helicoidal mode with tube voltage 120 kVp and automatically adjusted tube current. The other CT settings were: slice thickness 3.75 mm, spacing between slices 3.27 mm, in-plane inter-voxel spacing ≈1.37 mm in both directions, and image size 512 px × 512 px. All scans were performed on a Discovery 710 PET/CT system (GE Healthcare, Chicago, IL, USA).

2.3. Lesion Segmentation

Delineation of the region of interest (ROI) was performed by consensus by a panel of one radiation oncologist (I.P., >15 yr experience) and one nuclear medicine specialist (B.P., >20 yr experience) on the cross-platform, open-access software LIFEx 5.10 [34,35]. The process was carried out manually and separately for the CT and PET scans with no automatic transposition of the resulting ROI from PET to CT, or vice versa.

2.4. Extraction of Radiomics Features

From both the PET and CT images, we extracted five conventional features, four first-order statistics, six second-order statistics, and three shape features, as detailed in Table 2. For the features that require binning (marked with an asterisk in the Table), we used absolute quantisation into 256 levels over the following fixed-width intervals: [-1230 HU, 235 HU] for CT ("lung" window [36], giving a bin width of ≈5.7 HU) and [0 SUV, 40 SUV] for PET (bin width ≈0.16 SUV). Features from GLCM and NGTDM were computed using 26-connectivity and inter-voxel distance $\delta = 1$. No further pre-processing, like image normalisation, spatial resampling, or filtering was applied to the CT or PET images before feature extraction. Post-processing involved feature scaling through min–max normalisation ([37], [Equation 3.8]). The calculations were performed through the open-source Python

package Pyradiomics 2.2.0, to the documentation of which we refer the reader for details about feature definition, implementation, and mathematical formulae [38,39]. Preliminary conversion from DICOM (.dcm) format (in which the original scans were available) to Pyradiomics-readable NRRD format (.nrrd) was carried out on the open-source 3D Slicer platform 4.10.2 [40,41].

Table 2. Summary table of the radiomics features considered in the study. Key to abbreviations: 'GLCM' = Grey-Level Co-occurrence Matrix, 'NGTDM' = Neighbouring Gray-Tone Difference Matrix. Features that require binning are marked with an asterisk.

Group/Name	Definition/Interpretation
Conventional features	
Maximum	Maximum value (corresponds to SUV_{max} when computed on PET).
Minimum	Minimum value (corresponds to SUV_{min} when computed on PET).
Mean	Mean (average) value (corresponds to SUV_{mean} when computed on PET).
MaxAxialDiam	Largest pairwise Euclidean distance between the surface mesh of the lesion in the axial plane.
Volume	The volume of the lesion computed by summing up the volume of each voxel in the ROI. Usually referred to as Metabolic Tumour Volume (MTV) when computed on PET.
First-order statistics	
StdDev	Standard deviation, a measure of the data variability. Low values indicate the data are close to the mean; high values that they are spread over a large range.
Skewness	A measure of the distribution's departure from symmetry. Negative values indicate a left-skewed distribution (peak toward the right, left tail longer), positive values a right-skewed distribution (peak toward the left, right tail longer).
Kurtosis	A measure of the tailedness of the data compared to that of a normal distribution. High values indicate strongly-tailed data (presence of outliers), low values weakly-tailed data (absence of outliers).
Uniformity *	Sum of the squared probabilities of the intensity levels after binning. Measures the distribution's departure from uniformity. High values indicate that few intensity levels are more likely to occur than the other levels, low values that all the intensity levels are equally likely to occur.
Second-order statistics	
GLCM_DiffVar *	A measure of heterogeneity in which the occurrence probability of pairs of voxels whose intensity difference is far from the average is weighted more than that of pairs of pixels whose difference is close to the average.
GLCM_Energy *	The equivalent to first-order Uniformity (see above) for the joint distribution of pairs of voxel intensities. High values indicate that few pairs of intensity levels are more likely to occur than the other pairs and vice versa.
GLCM_Entropy *	A measure of the amount of information carried by the two-dimensional distribution of pairs of voxel intensities. High values indicates large variability/randomness, low values small variability/randomness.
NGTDM_Busyness *	A measure of the rate of change inversely weighted by the difference in magnitude between the intensity levels.
NGTDM_Coarseness *	A measure of the spatial rate of change between the intensity level of adjacent voxels. Can be interpreted as the size of the primitives in the image: higher values indicate lower spatial change therefore a locally more uniform texture.
NGTDM_Complexity *	A measure of the overall complexity of the image. It is related to the presence of primitive components in the image and the amount of rapid changes in the voxel intensities
Shape features	
Elongation	The squared inverse ratio between the largest and the second-largest principal components in the ROI shape. Values close to 0.0 indicate maximal elongation (a line-like, thin object), values close to 1.0 an object with approximately symmetric cross-section (like a square or a circle).
Flatness	The squared inverse ratio between the largest and smallest principal components in the ROI shape. Values close to 0.0 indicate a flat object, values close to 1.0 a sphere-like object.
Sphericity	A measure of the closeness of the ROI shape to that of a sphere. The value ranges between 0.0 and 1.0 with the latter indicating a perfectly spherical ROI.

2.5. Statistical Analysis

We carried out a statistical analysis in two steps. In the first, we tested each of the CT and PET features for significant differences between the benign and malignant groups. In the second, we determined whether shape and texture features could lead to a better prediction of malignant and benign nodules compared with conventional imaging features alone.

2.5.1. Significant Features

Hypothesis testing was based on non-parametric Wilcoxon-Mann-Witney test [42], the null hypothesis being that the distribution of each feature was the same in the benign and malignant groups. We set a significance level $\alpha = 0.05$ and applied Bonferroni's correction to protect from type-I error due to multiple tests [43]. Since we tested 18 features and two image modalities, differences were deemed significant if the *p*-value $< \alpha/(18 \times 2)$.

2.5.2. Assessment of Prediction Accuracy

To discriminate between benign and malignant nodules, we considered four feature combinations: two "base"' sets made up of conventional imaging features ("CT base", "PET base"), and two "enhanced" versions ("CT enhanced", "PET enhanced") in which the conventional features were complemented with texture and shape features. For all the sets, we only took into account those features that were significantly different between the benign and malignant groups as results from Tables 3 and 4 (see also Figure 2 for a round-up diagram).

To assess the ability of the above feature sets to discriminate between benign vs. malignant nodules, we built, for each set, three prediction models based on the following classification strategies: Classification Tree ("ClT"), k-Nearest Neighbours (KNN), and Naïve Bayes ("NB"). We used $k = 1$ and L_1 ("cityblock") distance for KNN, a Gaussian kernel for NB, and the default settings available in scikit.learn [44] for ClT.

For all the combination feature sets/classifiers, the ability to discriminate benign vs. malignant lesions was estimated following the same protocol described in [23]—that is, we used 80% of the samples of each class (train set) to train the prediction model, and the remaining 20% (test set) to estimate its accuracy. For a stable estimation, we repeated the random split into training and test sets 200 times and averaged the results. For each classifier, we checked for statistically significant differences in accuracy, sensitivity, and specificity between the "base" and "enhanced" sets. To this end, we again used Wilcoxon-Mann-Witney test ($\alpha = 0.05$) over the performance figures resulting from the 200 splits.

CT base (CTb)	CT enhanced (CTe)	PET base (PETb)	PET enhanced (PETe)
- Maximum - Mean - MaxAxialDiameter - Volume	- Maximum - Mean - MaxAxialDiameter - Volume - Skewness - GLCM_Energy - GLCM_Entropy - NGTDM_Coarseness	- Minimum (SUV_{min}) - Maximum (SUV_{max}) - Mean (SUV_{mean}) - MaxAxialDiameter - Volume (MTV)	- Minimum (SUV_{min}) - Maximum (SUV_{max}) - Mean (SUV_{mean}) - MaxAxialDiameter - Volume (MTV) - StdDev - Skewness - Uniformity - GLCM_DiffVar - GLCM_Energy - GLCM_Entropy - NGTDM_Busyness - NGTDM_Coarseness - NGTDM_Complexity - Flatness

Figure 2. The four feature sets used to build the prediction models.

3. Results

Tables 3 and 4 compare the values of the radiomics features in the benign ("N") and malignant ("P") groups. The CT features (Table 3) indicate that nodule size (volume, maximum diameter) and tissue density (mean, maximum) were significantly higher in the malignant group. The density distribution was also significantly more left-skewed in the malignant lesions than in the benign ones. The second-order statistics show that GLCM energy and entropy were, respectively, significantly lower and higher in the malignant group—both results indicating higher heterogeneity of the malignant lesions compared with the benign ones. This is also consistent with NGTDM coarseness, which was significantly higher in the benign nodules (more uniform texture) than in the malignant ones.

Table 3. Pairwise comparison between average feature values of CT radiomics features in the malignant ("P") and benign ("N") groups. Symbol "1" in the "Unit" column indicates a dimensionless feature.

Feature	Unit	N [Mean ± SD]	P [Mean ± SD]	*p*-Value	Alpha	Significant
Minimum	HU	−717.2 ± 104.3	−683.3 ± 150.9	0.0936	0.05/36	No
Maximum	HU	41.8 ± 218.7	64.5 ± 70.1	0.0005	0.05/36	Yes
Mean	HU	−243.0 ± 126.9	−155.2 ± 112.3	0.0001	0.05/36	Yes
StdDev	HU	173.75 ± 34.13	157.95 ± 42.89	0.9754	0.05/36	No
Skewness	1	−0.52 ± 0.62	−1.07 ± 0.74	0.0001	0.05/36	Yes
Kurtosis	1	2.96 ± 1.07	4.14 ± 3.38	0.0088	0.05/36	No
Uniformity	1	0.03 ± 0.01	0.03 ± 0.02	0.2966	0.05/36	No
GLCM_DiffVar	1	455.5 ± 197.6	391.2 ± 193.7	0.9520	0.05/36	No
GLCM_Energy	1	0.012 ± 0.013	0.005 ± 0.003	0.0003	0.05/36	Yes
GLCM_Entropy	1	7.339 ± 1.535	8.432 ± 0.934	<0.0001	0.05/36	Yes
NGTDM_Busyness	1	0.012 ± 0.006	0.012 ± 0.005	0.4791	0.05/36	No
NGTDM_Coarseness	1	0.021 ± 0.015	0.012 ± 0.010	<0.0001	0.05/36	Yes
NGTDM_Complexity	1	70,020.8 ± 41,222.8	76,640.8 ± 48,656.5	0.2455	0.05/36	No
VoxelVolume	mm^3	1820.8 ± 2083.8	3647.6 ± 3041.8	<0.0001	0.05/36	Yes
MaxAxialDiameter	mm	16.8 ± 6.3	21.9 ± 6.3	<0.0001	0.05/36	Yes
Sphericity	1	0.769 ± 0.122	0.770 ± 0.077	0.7447	0.05/36	No
Elongation	1	0.742 ± 0.157	0.790 ± 0.116	0.0695	0.05/36	No
Flatness	1	0.560 ± 0.178	0.648 ± 0.127	0.0050	0.05/36	No

Table 4. Pairwise comparison between average feature values of PET radiomics features in the malignant ("P") and benign ("N") group. Symbol "1" in the "Unit" column indicates a dimensionless feature.

Feature	Unit	N [Mean ± SD]	P [Mean ± SD]	*p*-Value	alpha	Significant
Minimum	SUV	0.9 ± 0.4	1.3 ± 0.5	<0.0001	0.05/36	Yes
Maximum	SUV	3.0 ± 3.2	7.9 ± 3.8	<0.0001	0.05/36	Yes
Mean	SUV	1.6 ± 1.1	3.5 ± 1.3	<0.0001	0.05/36	Yes
StdDev	SUV	0.49 ± 0.68	1.50 ± 0.85	<0.0001	0.05/36	Yes
Skewness	1	0.42 ± 0.42	0.70 ± 0.41	0.0003	0.05/36	Yes
Kurtosis	1	2.65 ± 0.76	2.97 ± 0.80	0.0027	0.05/36	No
Uniformity	1	0.27 ± 0.21	0.06 ± 0.06	<0.0001	0.05/36	Yes
GLCM_DiffVar	1	13.1 ± 34.4	44.4 ± 44.4	<0.0001	0.05/36	Yes
GLCM_Energy	1	0.147 ± 0.171	0.015 ± 0.025	<0.0001	0.05/36	Yes
GLCM_Entropy	1	4.197 ± 2.123	7.222 ± 1.485	<0.0001	0.05/36	Yes
NGTDM_Busyness	1	0.867 ± 1.386	0.164 ± 0.282	<0.0001	0.05/36	Yes
NGTDM_Coarseness	1	0.173 ± 0.171	0.041 ± 0.031	<0.0001	0.05/36	Yes
NGTDM_Complexity	1	1248.0 ± 3700.8	5267.9 ± 6399.6	<0.0001	0.05/36	Yes
VoxelVolume	mm^3	2107.7 ± 2409.2	5285.5 ± 3668.3	<0.0001	0.05/36	Yes
MaxAxialDiameter	mm	18.1 ± 5.7	23.4 ± 6.1	<0.0001	0.05/36	Yes
Sphericity	1	0.809 ± 0.110	0.822 ± 0.099	0.2455	0.05/36	No
Elongation	1	0.748 ± 0.117	0.775 ± 0.113	0.1338	0.05/36	No
Flatness	1	0.524 ± 0.213	0.642 ± 0.117	0.0007	0.05/36	Yes

The radiomics features from PET showed that radiotracer uptake (SUV$_{min}$, SUV$_{max}$, and SUV$_{mean}$) was higher in the malignant group, as one would reasonably expect. First-order statistics, like standard deviation and uniformity, were also different—higher and lower in the malignant group, respectively—both indicating that malignant lesions had a higher degree of heterogeneity/disorder. This trend was consistent with the second-order features, all of which were significantly different between the two groups; specifically, GLCM diffvar, GLCM entropy, and NGTDM complexity were higher in the malignant group, whereas GLCM energy, NGTDM busyness, and NGTDM coarseness were lower.

As for the shape features, only the flatness computed on PET was significantly different between the two groups and higher in the malignant one.

Table 5 reports the estimated performance (accuracy, sensitivity, and specificity) of the different combination feature sets/classifiers. As can be seen, adding texture features to the basic models always increased accuracy (3.4–11.2 pp gain for the CT-based models, 2.2–10.0 pp for the PET-based) and sensitivity (3.4–12.8 pp for the CT-based models, 3.4–17.7 pp for the PET-based) by an appreciable margin. Specificity also increased for all the CT-based models (6.0–8.3 pp), but in one case, decreased for the PET-based models (-3.9–4.5 pp).

Table 5. Estimated performance of the different combination feature sets/classifiers and pairwise differences. Key to symbols: "ACC" = accuracy, "SP" = specificity, "SN" = sensitivity. Values are in %, differences in percentage points. Boldface figures indicate significant differences. For a comparison: accuracy of a random classifier (blind to prior class probabilities) = 50%; with prior class probabilities = 54.4%.

Model	Classifier								
	ClT			KNN			NBGaussian		
	ACC	SN	SP	ACC	SN	SP	ACC	SN	SP
CTb	58.8	67.6	42.2	59.5	65.9	47.5	69.5	83.5	43.3
CTe	62.2	69.6	48.2	70.7	78.7	55.8	74.3	86.9	50.8
PETb	73.4	78.7	63.5	72.5	76.9	64.4	72.2	69.9	76.6
PETe	75.7	82.1	63.6	77.1	81.4	69.1	82.4	87.6	72.6
CTb+PETb	71.2	77.1	60.1	74.4	78.9	66.1	70.6	71.3	69.1
CTe+PETe	72.3	78.9	59.8	73.7	75.7	70.1	80.4	88.5	65.2
CTe-CTb	**+3.4**	**+2.0**	**+6.0**	**+11.2**	**+12.8**	**+8.3**	**+4.8**	**+3.4**	**+7.5**
PETe-PETb	**+2.2**	**+3.4**	+0.1	**+4.6**	**+4.5**	**+4.7**	**+10.2**	**+17.7**	**−3.9**

4. Discussion

The recent literature has consistently emphasized the potentially useful role that shape and texture features from PET/CT could play in the characterisation of suspicious pulmonary nodules. Still, the validity and implications of these results need to be understood better before SPN radiomics can be translated into clinical practice [13,16].

One crucial point is to quantify the actual benefit that radiomics features can provide beyond conventional imaging parameters alone. In our experiments, shape and texture features were able to improve the overall accuracy of the prediction models by 2.2–11.2 pp. The increase in CT accuracy found here is comparable, in magnitude, with that reported by Wu et al. [25] and Balagurunathan et al. [23]. However, our overall accuracy was, in absolute terms, lower than that reported in the above references, most probably because those results were obtained with standard CT scans, ours with low-dose ones. For a comparison, Miwa et al. [32] obtained 62.9% accuracy with CT features extracted from low-dose scans, a result very much in line with our figures.

Texture analysis at CT (Table 3) showed that malignant lesions had higher GLCM entropy, lower GLCM energy, and lower NGTDM coarseness than benign lesions, all indicating that malignant nodules had a higher degree of heterogeneity than benign ones. This finding is in agreement with

the results reported by Zhao et al. [45] (though their study was two-dimensional), and in general, seems to confirm image heterogeneity at CT as an indicator of malignancy, as previous studies have suggested [46,47]. Skewness was also different between the benign and malignant groups, with the former exhibiting more negatively skewed distributions (excess of high vales) than the benign ones. The meaning and potential implications of this finding, however, are unclear and should be further evaluated in future studies.

It has long been speculated that some characteristics of pulmonary nodules at CT could help differentiate between benign and malignant lung nodules. In particular, irregular, "spiculated" lesions are usually considered more likely to be malignant than round ones with well-defined, smooth margins [1,11]. Although this hypothesis has found some confirmation in previous studies [20,48], we did not find further support for it in our investigation: among the shape features considered here (i.e., sphericity, elongation, and flatness), no one was significantly different between the benign and malignant groups.

The role of tissue density has also been a subject of debate in a number of studies and never clarified completely. Fat density (attenuation values between -120 HU and -40 HU) is considered a possible indicator of hamartomas, and attenuation values greater than 200 HU of calcification [10,11]. In [49], however, the authors concluded that baseline nodule density could not be used to discriminate between benign and malignant SPN. Here, we found that tissue density was significantly higher in the malignant group, although the intra-class variation was considerable (mean value -243.2 ± 126.9 HU and -155.2 ± 112.3 HU, respectively, for the benign and malignant nodules).

Fifteen PET features (Table 4) were significantly different between the benign and malignant groups—that is, nearly twice as many as the CT features. Besides, the models based on PET features in general outperformed those based on CT features, which is in agreement with previously published results [29,32,50]. This is interesting, since it suggests that assessing the heterogeneity of the radiotracer uptake via texture features from PET (as was discussed, for instance, in [5,6]) can be at least as important as doing the same on tissue density via texture features from CT, although the literature has mostly focused on the latter so far.

All the first-order and second-order statistics indicate that malignant lesions had higher uptake heterogeneity than the benign ones. In particular, standard deviation, GLCM diffvar, GLCM entropy, and GLCM complexity were higher for the positive group, whereas uniformity, GLCM energy, and GLCM coarseness were lower. NGTDM busyness was lower in the positive group, which is consistent with the larger uptake range in the malignant lesions, and is in agreement with the results reported in [28]. The uptake distribution in the malignant group was more positively skewed (excess of low vales) than in the benign one, although this finding is not readily interpretable and the potential implications of it are unclear.

Interestingly, the results showed some dependence on the classification model used. Whereas adding shape and texture features always increased the overall prediction accuracy, we noticed that with the Naïve Bayes classifier and CT features, there was a negative effect on sensitivity, which dropped by 3.0 pp.

5. Conclusions

We found that shape and texture features from ^{18}F-FDG PET/CT can provide added value in the discrimination between benign and malignant lung nodules compared with conventional imaging features alone. This work is not exempt from limitations—among them are the relatively contained sample size and the retrospective nature. The findings should be further confirmed in larger, ideally prospective studies. By design, we also focused on evaluating the potential benefit of radiomics features compared with conventional imaging features; therefore, the effects of other potentially relevant data, such as nodule location, previous exposure to risk factors, age and/or gender were not investigated here.

References

1. Ost, D.; Fein, A.; Feinsilver, S. The solitary pulmonary nodule. *N. Engl. J. Med.* **2003**, *348*, 2535–2542. [CrossRef] [PubMed]

2. Larici, A.; Farchione, A.; Franchi, P.; Ciliberto, M.; Cicchetti, G.; Calandriello, L.; del Ciello, A.; Bonomo, L. Lung nodules: Size still matters. *Eur. Respir. Rev.* **2017**, *26*, 170025. [CrossRef] [PubMed]

3. Tang, K.; Wang, L.; Lin, J.; Zheng, X.; Wu, Y. The value of 18F-FDG PET/CT in the diagnosis of different size of solitary pulmonary nodules. *Medicine* **2019**, *11*, e14813. [CrossRef]

4. Gould, M.; Tang, T.; Liu, I.L.; Lee, J.; Zheng, C.; Danforth, K.; Kosco, A.; Di Fiore, J.; Suh, D. Recent trends in the identification of incidental pulmonary nodules. *Am. J. Respir. Crit. Care Med.* **2015**, *192*, 1208–1214. [CrossRef]

5. García Vicente, A.; Pérez-García, V.; Soriano Castrejón, A. 18F-fluorodeoxyglucose positron emission tomography/computed tomography characterization of solitary pulmonary nodules: Can we do better? *J. Thorac. Dis.* **2015**, *8*, E215–E218.

6. Schillaci, O.; Calabria, F. Comments on characterization of solitary pulmonary nodules with 18F-FDG PET/CT relative activity distribution analysis. *J. Thorac. Dis.* **2015**, *10*, 1708–1712.

7. Nasim, F.; Ost, D. Management of the solitary pulmonary nodule. *Curr. Opin. Pulm. Med.* **2019**, *25*, 344–353. [CrossRef]

8. Evangelista, L.; Cuocolo, A.; Pace, L.; Mansi, L.; Del Vecchio, S.; Miletto, P.; Sanfilippo, S.; Pellegrino, S.; Guerra, L.; Pepe, G.; et al. Performance of FDG-PET/CT in solitary pulmonary nodule based on pre-test likelihood of malignancy: Results from the ITALIAN retrospective multicenter trial. *Eur. J. Nucl. Med. Mol. Imaging* **2018**, *11*, 1898–1907. [CrossRef]

9. Chan, E.; Gaur, P.; Ge, Y.; Kopas, L.; Santacruz, J.; Gupta, N.; Munden, R.; Cagle, P.; Kim, M. Management of the solitary pulmonary nodule. *Arch. Pathol. Lab. Med.* **2017**, *141*, 927–931. [CrossRef] [PubMed]

10. Edey, A.; Hansell, D. Incidentally detected small pulmonary nodules on CT. *Clin. Radiol.* **2009**, *64*, 872–884. [CrossRef]

11. Truong, M.; Ko, J.; Rossi, S.; Rossi, I.; Viswanathan, C.; Bruzzi, J.; Marom, E.; Erasmus, J. Update in the evaluation of the solitary pulmonary nodule. *Radiographics* **2014**, *34*, 1658–1679. [CrossRef] [PubMed]

12. MacMahon, H.; Naidich, D.; Goo, J.; Lee, K.; Leung, A.; Mayo, J.; Mehta, A.; Ohno, Y.; Powell, C.; Prokop, M.; et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. *Radiology* **2017**, *284*, 228–243. [CrossRef] [PubMed]

13. Bianconi, F.; Palumbo, I.; Spanu, A.; Nuvoli, S.; Fravolini, M.; Palumbo, B. PET/CT radiomics in lung cancer: An overview. *Appl. Sci.* **2020**, *5*, 1718. [CrossRef]

14. Kaseda, K. Recent and Current Advances in FDG-PET Imaging within the Field of Clinical Oncology in NSCLC: A Review of the Literature. *Diagnostics* **2020**, *10*, 561. [CrossRef] [PubMed]

15. Vallières, M.; Zwanenburg, A.; Badic, B.; Le Rest, C.; Visvikis, D.; Hatt, M. Responsible radiomics research for faster clinical translation. *J. Nucl. Med.* **2018**, *59*, 189–193. [CrossRef]

16. Wilson, R.; Devaraj, A. Radiomics of pulmonary nodules and lung cancer. *Transl. Lung Cancer Res.* **2017**, *6*, 86–91. [CrossRef]

17. Cook, G.; Goh, V. A Role for FDG PET Radiomics in Personalized Medicine? *Semin. Nucl. Med.* **2020**, In press. [CrossRef]

18. Mayerhoefer, M.; Materka, A.; Langs, G.; Häggström, I.; Szczypiński, P.; Gibbs, P.; Cook, G. Introduction to Radiomics. *J. Nucl. Med.* **2020**, *61*, 488–495. [CrossRef]

19. Krewer, H.; Geiger, B.; Hall, L.; Goldgof, D.; Gu, Y.; Tockman, M.; Gillies, R. Effect of texture features in computer aided diagnosis of pulmonary nodules in low-dose computed tomography. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 3887–3891. Article number 6722416.

20. Dhara, A.; Mukhopadhyay, S.; Dutta, A.; Garg, M.; Khandelwal, N. A Combination of Shape and Texture Features for Classification of Pulmonary Nodules in Lung CT Images. *J. Digit. Imaging* **2016**, *29*, 466–475. [CrossRef]

21. Hawkins, S.; Wang, H.; Liu, Y.; Garcia, A.; Stringfield, O.; Krewer, H.; Li, Q.; Cherezov, D.; Gatenby, R.; Balagurunathan, Y.; et al. Predicting Malignant Nodules from Screening CT Scans. *J. Thorac. Oncol.* **2016**, *11*, 2120–2128. [CrossRef]

22. Ferreira, J.J.; Oliveira, M.; de Azevedo-Marques, P. Characterization of Pulmonary Nodules Based on Features of margin Sharpness and Texture. *J. Digit. Imaging* **2018**, *31*, 451–463. [CrossRef]

23. Balagurunathan, Y.; Schabath, M.; Wang, H.; Liu, Y.; Gillies, R. Quantitative Imaging features Improve Discrimination of Malignancy in Pulmonary nodules. *Sci. Rep.* **2019**, *9*, 8528. [CrossRef]

24. Gao, N.; Tian, S.; Li, X.; Huang, J.; Wang, J.; Chen, S.; Ma, Y.; Liu, X.; Guo, X. Three-Dimensional Texture Feature Analysis of Pulmonary Nodules in CT Images: Lung Cancer Predictive Models Based on Support Vector Machine Classifier. *J. Digit. Imaging* **2019**, *33*, 414–422. [CrossRef] [PubMed]

25. Wu, W.; Pierce, L.; Zhang, Y.; Pipavath, S.; Randolph, T.; Lastwika, K.; Lampe, P.; Houghton, A.; Liu, H.; Xia, L.; et al. Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: A case-control study. *Eur. Radiol.* **2019**, *29*, 6100–6108. [CrossRef] [PubMed]

26. Wu, W.; Hu, H.; Gong, J.; Li, X.; Huang, G.; Nie, S. Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis. *Phys. Med. Biol.* **2019**, *64*, 035017. [CrossRef] [PubMed]

27. Wang, L.; Chen, Y.; Tang, K.; Lin, J.; Zhang, H. The Value of 18F-FDG PET/CT Mathematical Prediction Model in Diagnosis of Solitary Pulmonary Nodules. *Biomed. Res. Int.* **2018**, 9453967. [CrossRef] [PubMed]

28. Chen, S.; Harmon, S.; Perk, T.; Li, X.; Chen, M.; Li, Y.; Jeraj, R. Using neighborhood gray tone difference matrix texture features on dual time point PET/CT images to differentiate malignant from benign FDG-avid solitary pulmonary nodules. *Cancer Imaging* **2019**, *19*, 56. [CrossRef]

29. Taralli, S.; Scolozzi, V.; Foti, M.; Ricciardi, S.; Forcione, A.; Cardillo, G.; Calcagni, M. 18F-FDG PET/CT diagnostic performance in solitary and multiple pulmonary nodules detected in patients with previous cancer history: Reports of 182 nodules. *Eur. J. Nucl. Med. Mol. Imaging* **2019**, *46*, 429–436. [CrossRef]

30. Karyağar, S.; Koç, Z.; Karyağar, S.; Bekar, Y. Diagnostic performance of 18F-FDG PET/CT in solitary pulmonary nodules of non-smokers. *Turk. J. Thorac. Cardiovasc. Surg.* **2017**, *25*, 235–241. [CrossRef]

31. Wang, C.; Guan, Y.; Zuo, C.; Hua, F.; Chen, L.; Zhang, W.; Pan, Y.; Yuan, Y.; Zhou, Y. Value of the texture feature for solitary pulmonary nodules and mass lesions based on PET/CT. In Proceedings of the 4th International Conference on Bioinformatics and Biomedical Engineering, ICBBE, Chengdu, China, 18–20 June 2010; p. 5514710.

32. Miwa, K.; Inubushi, M.; Wagatsuma, K.; Nagao, M.; Murata, T.; Koyama, M.; Koizumi, M.; Sasaki, M. FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur. J. Radiol.* **2014**, *83*, 715–719. [CrossRef]

33. Jia, Y.; Gong, W.; Zhang, Z.; Tu, G.; Li, J.; Xiong, F.; Hou, H.; Zhang, Y.; Wu, M.; Zhang, L. Comparing the diagnostic value of 18F-FDG-PET/CT versus CT for differentiating benign and malignant solitary pulmonary nodules: A meta-analysis. *J. Thorac. Dis.* **2019**, *11*, 2082–2098. [CrossRef] [PubMed]

34. LIFEx Soft. Available online: http://www.lifexsoft.org (accessed on 2 April 2020).

35. Nioche, C.; Orlhac, F.; Boughdad, S.; Reuze, S.; Goya-Outi, J.; Robert, C.; Pellot-Barakat, C.; Soussan, M.; Frouin, F.; Buvat, I. LIFEx: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* **2018**, *78*, 4786–4789. [CrossRef] [PubMed]

36. Broder, J. Chapter Imaging of Nontraumatic Abdominal Conditions. In *Diagnostic Imaging for the Emergency Physician;* Elsevier Saunders: Philadelphia, PA, USA, 2011; pp. 445–577.

37. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2012.

38. van Griethuysenand, J.; Fedorovand, A.; Parmarand, C.; Hosnyand, A.; Vivek Narayanand, A.; Beets-Tanand, R.; Fillion-Robinand, J.C.; Pieperand, S.; Aerts, H. Computational Radiomics System to decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef] [PubMed]

39. Py-Radiomics: Open-Source Radiomics Library Written in Python. Available online: https://www.radiomics.io/pyradiomics.html (accessed on 26 February 2020).

40. Kikinis, R.; Pieper, S.; Vosburgh, K. *Intraoperative Imaging and Image-Guided Therapy*; Chapter 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support; Springer: New York, NY, USA, 2013; pp. 277–289.

41. 3DSlicer: A Multi-Platform, Free, Open Source Software Package For Visualization and Medical Image Computing. Available online https://www.slicer.org/ (accessed on 21 April 2020).

42. Kanji, G. *100 Statistical Tests*, 3rd ed.; SAGE Publications: London, UK, 2006.

43. Chalkidou, A.; O'Doherty, M.; Marsden, P. False discovery rates in PET and CT studies with texture features: A systematic review. *PLoS ONE* **2015**, *10*, e0124165. [CrossRef]

44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

45. Zhao, Q.; Shi, C.Z.; Luo, L.P. Role of the texture features of images in the diagnosis of solitary pulmonary nodules in different sizes. *Chin. J. Cancer Res.* **2014**, *26*, 451–458.

46. Ganeshan, B.; Goh, V.; Mandeville, H.; Ng, Q.; Hoskin, P.; Miles, K. Non-small cell lung cancer: Histopathologic correlates for texture parameters at CT. *Radiology* **2013**, *266*, 326–336. [CrossRef]

47. Ganeshan, B.; Miles, K. Quantifying tumour heterogeneity with CT. *Cancer Imaging* **2013**, *13*, 140–149. [CrossRef]

48. Iwano, S.; Nakamura, T.; Kamioka, Y.; Ishigaki, T. Computer-aided diagnosis: A shape classification of pulmonary nodules imaged by high-resolution CT. *Comput. Med Imaging Graph.* **2005**, *29*, 565–570. [CrossRef]

49. Xu, D.; van Klaveren, R.; de Bock, G.; Leusveld, A.; Dorrius, M.; Zhao, Y.; Wang, Y.; de Koning, H.; Scholten, E.; Verschakelen, J.; et al. Role of baseline nodule density and changes in density and nodule features in the discrimination between benign and malignant solid indeterminate pulmonary nodules. *Eur. J. Radiol.* **2009**, *70*, 492–498. [CrossRef]

50. Christensen, J.; Nathan, M.; Mullan, B.; Hartman, T.; Swensen, S.; Lowe, V. Characterization of the solitary pulmonary nodule: 18F-FDG PET versus nodule-enhancement CT. *Am. J. Roentgenol.* **2006**, *187*, 1361–1367. [CrossRef] [PubMed]