European Journal of Radiology 121 (2019) 108735

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Potential feature exploration and model development based on 18F-FDG PET/CT images for differentiating benign and malignant lung lesions

Ruiping Zhang[a,1], Lei Zhu[b,1], Zhengting Cai[c], Wei Jiang[b], Jian Li[d], Chengwen Yang[a], Chunxu Yu[e], Bo Jiang[a], Wei Wang[a], Wengui Xu[b], Xiangfei Chai[c], Xiaodong Zhang[f], Yong Tang[d,*]

[a] Department of Radiation Oncology, Department of Radiation Physics, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Huanhu West Road, Hexi District, Tianjin, 300060, China
[b] Department of Nuclear Medicine, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Huanhu West Road, Hexi District, Tianjin, 300060, China
[c] Huiying Medical Technology Co., Ltd., Dongcheng Science and Technology Park, Dongcheng Road, Haidian District, Beijing, 100192, China
[d] Department of Pancreas Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Huanhu West Road, Hexi District, Tianjin, 300060, China
[e] Department of Physics, Nankai University, Weijin Road, Nankai District, Tianjin, 300071, China
[f] Department of Radiation Oncology, Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, 1840 Old Spanish Trail, Houston, TX, 77054, USA

ARTICLE INFO

Keywords:
Lung lesion
CT-radiomics features
PET metabolic parameters
Potential feature

ABSTRACT

Purpose: The study is to explore potential features and develop classification models for distinguishing benign and malignant lung lesions based on CT-radiomics features and PET metabolic parameters extracted from PET/CT images.

Materials and methods: A retrospective study was conducted in baseline 18 F-flurodeoxyglucose positron emission tomography/ computed tomography (18 F-FDG PET/CT) images of 135 patients. The dataset was utilized for feature extraction of CT-radiomics features and PET metabolic parameters based on volume of interest, then went through feature selection and model development with strategy of five-fold cross-validation. Specifically, model development used support vector machine, PET metabolic parameters selection used Akaike's information criterion, and CT-radiomics were reduced by the least absolute shrinkage and selection operator method then forward selection approach. The diagnostic performances of CT-radiomics, PET metabolic parameters and combination of both were illustrated by receiver operating characteristic (ROC) curves, and compared by Delong test. Five groups of selected PET metabolic parameters and CT-radiomics were counted, and potential features were found and analyzed with Mann-Whitney *U* test.

Results: The CT-radiomics, PET metabolic parameters, and combination of both among five subsets showed mean area under the curve (AUC) of 0.820 ± 0.053, 0.874 ± 0.081, and 0.887 ± 0.046, respectively. No significant differences in ROC among models were observed through pairwise comparison in each fold (P-value from 0.09 to 0.81, Delong test). The potential features were found to be SurfaceVolumeRatio and SUVpeak ($P < 0.001$ of both, *U* test).

Conclusion: The classification models developed by CT-radiomics features and PET metabolic parameters based on PET/CT images have substantial diagnostic capacity on lung lesions.

Abbreviations: 18F-FDGPET/CT, 18F-flurodeoxyglucose positron emission tomography/ computed tomography; AIC, Akaike's information criterion; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the curve; SVM, support vector machine; OSEM, ordered subset expectation maximization; VOI, volume of interest; SUV, standardized uptake volume; MTV, metabolic tumor volume

* Corresponding author.
E-mail address: tli001@163.com (Y. Tang).

[1] Ruiping Zhang and Lei Zhu contributed equally to this work. They worked together to select clinical data, analyze the results and solve problems encountered in this study and wrote the manuscripts.

https://doi.org/10.1016/j.ejrad.2019.108735
Received 15 July 2019; Received in revised form 11 September 2019; Accepted 31 October 2019

1. Introduction

Lung cancer has been one of the most common forms of cancers and a leading cause of mortality in the world. Early diagnosis of lung cancer through screening has become more important in clinical practice [1,2], and accurate classification of lung lesions is a prerequisite for adequate clinical management.

As a functional imaging technique, 18 F-fluodeoxyglucose positron emission tomography (18 F-FDG PET) measuring metabolic activity has been used for assessment and diagnosis of lung lesions [3], which is frequently based on the quantitative parameters named standardized uptake values of a tracer (SUV) [4]. However, PET scans usually were accompanied with side effects, such as high false positive rates on granulomatous disease and inflammatory pseudotumor, financial cost and radiation exposure. The revolutionized combination of positron emission tomography and computed tomography (PET/CT) has greatly increased the strong ability to accurately diagnose properties of lung lesions, even assisting in the staging of lung cancer [5–7]. Shanna et al. [7] revealed that PET/CT combining the strength of anatomic and metabolic imaging demonstrated an amazing performance in classifying solitary pulmonary nodules as benign or malignant. Yet the interpretation of imaging in their study was mainly performed by radiologists and nuclear medicine physicians, therefore the relevant results were subject to individual physician experience which results in more random and not conducive to grassroots promotion. Meanwhile, the above-mentioned inherent defects of PET have not been well resolved, and the less expensive and more popular modality of CT did not show satisfactory consequences in the study with low accuracy and low specificity. Consequently, more stable and effective features from CT images should be added.

Recently, radiomics analysis has attracted increased attention as a potential classification tool and is showing promising prospects to improve the diagnostic accuracy of imaging on CT [8–11].

Radiomics analysis is the process of conversion of image information into high-dimensional and mineable data via high-throughput extraction of sets of quantitative features [12]. Choi et al. [13] have confirmed than an SVM(support vector machine)-LASSO (least absolute shrinkage and selection operator) model developed with two CT radiomics features achieved an accuracy of 84.6% for diagnosis of lung cancer. Hawkins et al. [14] built and tested classifiers for successfully predicting malignant nodules.

Mediastinal lymph nodes characterization was also studied by Digumarthy et al. [15] based on radiomics features. The combination of PET metabolic parameters with radiomics features extracted from PET images was reported for prognosis of recurrence and survival in lung cancer [16]. The performance of the combination of radiomics features from CT image components and PET metabolic parameters for the diagnosis of lung lesions has not been investigated to date.

In our study, CT-radiomics features and PET metabolic parameters were extracted from only the CT image components and PET images of 18 F-FDG PET/CT respectively. Three models (CT-radiomics features, PET metabolic parameters, and a combination of both) were developed to differentiate malignant and benign lung lesions. Moreover, CT-radiomics features and PET metabolic parameters were further explored to obtain potential features representing certain lung lesion clinical characteristics.

2. Materials and methods

2.1. Patients

The study was approved by the Tianjin Cancer Hospital Medical Ethics Committee and informed consent was waived. The study enrolled patients who had undergone 18 F-FDG PET/CT scanning in our hospital from January 2014 to January 2016, for whom a pathology diagnosis was available. The inclusion criteria for patients were well-defined pathology after biopsy or surgery. Exclusion criteria for patients included the following: 1) radiation therapy or chemotherapy prior to PET/CT scanning, 2) the presence of calcium, cavity, air bubble in the lesion, bronchoalveolar pneumonia, peripheral consolidation, and bubbles in mediastinum, 3) the lung lesion volume less than 1cm^3, and 4) the lesion invading into hilum of lung, mediastinum, and pericardium.

2.2. Image acquisition and reconstruction

The 18 F-FDG PET/CT scanning was performed with a hybrid PET/CT scanner (Discovery ST4, GE Healthcare, Chicago, IL, USA). All patients fasted for at least 6 h and were injected with 18 F-FDG (3.70–4.81 m Bq /kg) depending on body weight at 66 ± 10 min before data acquisition. The protocol included an initial CT scan followed by PET acquisition. PET data were obtained in three-dimensional mode using 6 to 8 bed positions with 2 min per bed position. PET images were reconstructed with an attenuation correction calculated from registered CT images using the ordered subset expectation maximization (OSEM) iterative algorithm [17]. The initial CT image acquisition was conducted with slice thickness 3.75 mm (120KVp, 100 mA), and reconstructed to a 512 × 512 matrix (voxels size, 0.98 × 0.98 × 3.27 mm^3). Then PET images, CT images and fused PET/CT data were available for viewing in coronal, sagittal and axial planes using an Xeleris review station (GE Healthcare).

2.3. PET metabolic parameters extraction

We chose nine metabolic parameters of 18 F-FDG uptake that are generally used for measurement of intratumoral heterogeneity. They were delineated using the PET VCAR visualization product (GE Healthcare) on a workstation (Advantage version 4.6, GE Healthcare) by cutoff of 40% maximum standardized uptake value (SUVmax). The nine PET metabolic parameters are as follows: MTV, SUVmax, SUVmean, SUVpeak, SULmax, SULmean, SULpeak, SUVTLG and SULTLG, and the detailed description of these PET metabolic parameters was shown in Appendix A in Supplementary materials.

2.4. Lung lesion segmentation and radiomics feature extraction on CT image

The volume of interest (VOI) for extracting radiomics feature was contoured on the CT image component of the 18 F-FDG PET/CT scans by two physicians, each with 10 years of experience. Attention was paid to avoiding adjacent vascular structures, chest wall or mediastinum [18]. Each lesion was delineated on every sequential slice.

In order to remove individual acquisition differences and anisotropy, the images were normalized and resampled into the voxel size of 1 × 1 × 1mm^3 on a Radcloud platform (version 2.1.2, http://radcloud.cn/, Huiying Medical Technology Co., Ltd, Beijing, China) before all of work of radiomics feature extraction, then the total of 1029 radiomics feature of VOI from CT images were also extracted on platform.

The radiomics features were divided into three groups: 1) first order features, 2) shape features, 3) texture features group. In addition, texture features were divided into the following three subgroups: Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM) and Gray Level Size Zone Matrix (GLSZM). Each specific feature in the groups of first-order features, shape features, GLCM, GLRLM and GLSZM was listed in Table A1 of Appendix A in Supplementary materials.

Only shape features were calculated on the original image, while the other groups (1 and 3) were computed on both the original image and derived images. The images were processed using filters including the exponential filter, square filter, square root filter, logarithm filter and wavelet decomposition filter [19,20].

Besides, the wavelet decomposition includes the following three-dimensional wavelet transforms by directional low-pass (L) and high-

pass (H) filtering: wave-LHL, wave-LHH, wave-HLL, wave-LLH, wave-HLH, wave-HHH, wave-HHL and wave-LLL.

In conclusion, the feature distribution: the number of all features (1029) = shape (15) + [first-order features (19) + GLCM (27) + GLRLM (16) + GLSZM (16)] X the numbers of original and processed images (13). Detailed information is listed in Appendix A in Supplementary materials [19].

2.5. Feature selection and model development

To develop a robust and stable diagnosis model, a five-fold cross-validation method was conducted based on our data set: The whole set was randomly partitioned into five equal sized subsets, of which one subset was utilized as the test set and the remaining four subsets as the training set, repeating five times to ensure that each of five subsets was used as the test set only once. Each time, the training set went through three procedures of feature standardization, feature selection, and model development, then features in the test set were also standardized and selected based on the relevant statistics stored by the training set to evaluate the model. All of the above work was done using Python3.6 (https://www.python.org/) in an Anaconda3 platform (https://www.anaconda.com) with scikit-learn (https://scikit-learn.org/) and matplotlib packages (https://matplotlib.org/), and the parameters of methods not specifically mentioned in the latter part were used by default in scikit-learn package. The workflow chart of feature selection and model development is displayed in Fig. 1.

2.5.1. Feature standardization

In order to avoid some features overwhelming others, each feature (including CT-radiomics features and PET metabolic parameters) was standardized independently by removing the mean and scaling to unit variance in the training set, as follows:

$$X_{ij}^{'} = \frac{X_{ij} - \mu_i}{\sigma_i}$$

where i is the i^{th} feature, j is the j^{th} patient, $X_{ij}^{'}$ is the standardized feature value, X_{ij} is the original feature value, and μ_i *and* σ_i are the original feature's mean value and standard deviation respectively.

2.5.2. Feature selection

Feature selection is a general requirement for many machine learning studies. It can help to improve models' performance ability and decrease dataset redundancy. Two kinds of strategies were used to select PET metabolic parameters and CT-radiomics features respectively.

2.5.2.1. PET metabolism parameters. An inner five-fold cross-validation using complete measure (exhausting all possible combinations of PET metabolism parameters) [21] with the support vector machine (SVM) [22] method in the training set was performed to search for the optimal parameter subset according to the minimum of mean Akaike's information criterion (AIC) [23] (mean AIC values of all possible subsets in each group shown in Appendix B in Supplementary materials). AIC follows the 'principle of parsimony', inclined to choose the subset with fewer parameters with guarantee of a satisfactory performance. The used formula of AIC is as follows:

$$AIC = n\ln\left(\frac{RSS}{n}\right) + 2m$$

where n is the number of samples, m is the number of features, and RSS is residual sum of squares.

2.5.2.2. CT-radiomics features. The least absolute shrinkage and selection operator (LASSO) method [24], which is suitable for the regression of high-dimension data with a highly redundant feature space, was used to select the most useful diagnostic feature from the training set. The optimal α, which is the coefficient of regularization in the LASSO method, was selected using inner ten-fold cross-validation in the training set via minimum average mean square error (MSE). Subsequently, the radiomics parameters with non-zero coefficients in the LASSO model generated by the whole training set with the optimal α were selected.

To avoid over-fitting caused by above feature selection process, the number of selected radiomics parameters was further reduced using a forward selection approach (starting with no features in the model, and successively adding the features ranked by the absolute values of coefficients in the LASSO model to improve the model fitting the criterion) [25] in a five-fold cross-validation loop with the SVM method. The criterion of confirming a final number in our study was the least

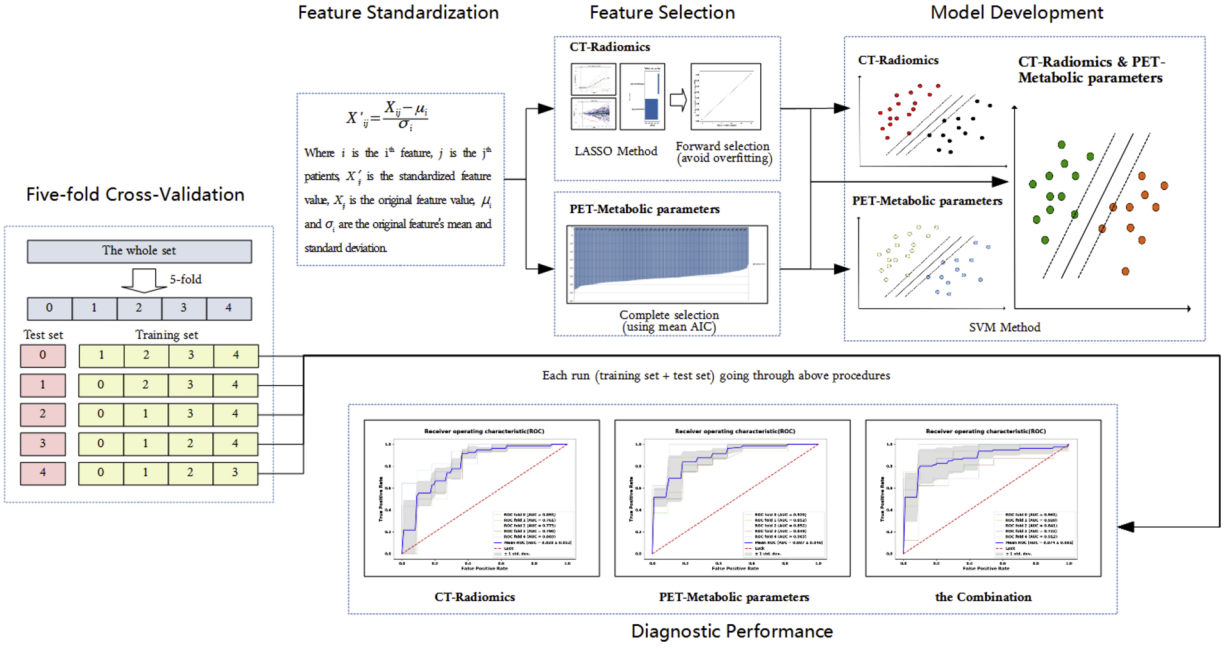

Fig. 1. The workflow chart of feature selection and model development using five-fold cross-validation with support machine vector (SVM) method. The models were developed by the feature selection in the training set and evaluated by the test set in each run.

one with the mean area under the curve (AUC) over 0.8, or the most one if the mean AUC of each possible number was unable to reach 0.8.

2.5.3. Model development

Three diagnostic models of CT-radiomics features, PET metabolism parameters and the combination of both were also developed by the SVM method in the training set based on the selected PET metabolic parameters and CT-radiomics features.

Support vector machine is an effective and efficient supervised learning method, especially in limited sample sizes and high dimensional spaces, which uses several of training samples (called support vectors) to build the decision function based on specific kernel function, and the penalty parameter C to avoid over-fitting [26]. In our study, the hyper-parameters of methods followed default settings of packages, using the radial basis Function (RBF) kernel, and the value of C setting 1.

2.6. Statistical analysis

The diagnostic performance of the three models was evaluated by the independent test set in each run, illustrated by the receiver operating characteristic curve (ROC curve) with indices of AUC, accuracy, sensitivity and specificity. The criterion value dividing the benign and malignant predictive values was determined by the maximum of the Youden index [27] (calculating the sum of sensitivity and specificity and then subtracting one over all possible threshold values), which also corresponded with the point on the ROC curve farthest from the diagonal line named 'Luck' in plots. The significant differences in the ROC curves between models were compared using the Delong test [28] (Medcalc version 15.2.2, http://www.medcalc.org).

Five groups of the selected PET metabolic parameters and CT-radiomics features were counted, and the features appearing five times (regarded as 'potential features' by us) were analyzed statistically. The Mann-Whitney U test [29] was performed to assess the differences of potential features between benign and malignant groups. We also plotted histograms of potential features of the two groups and visualized the distribution density estimated by the Gaussian kernel with appropriate bandwidth.

For patient clinical characteristics, continuous variables (e.g., age) were presented as the mean standard deviation, while categorical variables (e.g., pathology type and gender) were presented as the percentages. A P-value less than 0.05 (two tailed) was considered as statistically significant in our study.

3. Results

3.1. Patient characteristics

One hundred and thirty-five patients were included in our study (age: 58 ± 11years). Of these 87(64%) were men and 48 (36%) were women, and 40% were diagnosed as benign and 60% malignant based on clinical pathology results. Detail clinical information of patients is summarized in Table 1.

3.2. Features selection

Of nine PET metabolism parameters, only the subset including one parameter named SUVpeak was selected. For selection of CT-radiomics features, five groups of significant features were selected by the LASSO method and forward selection approach (all of the selected radiomics features are shown in Table 2 and Fig. 2, and the details of the CT-radiomics feature selection are shown in Appendix C in Supplementary materials). Ten of these features appeared at least once in five runs of cross-validation; the feature named 'RootMeanSquared' after logarithm filtering appeared three times, and the 'SurfaceVolumeRatio' from the shape feature group appeared five times.

Table 1
Clinical information of patients with lung lesions.

Lesion Pathology	Number (Percentage)
Granulomatous Lesion	23 (17.0%)
Inflammatory Pseudotumor	29 (21.5%)
Hemangioma	1 (0.7%)
Fibrocellullar Tumor	1 (0.7%)
Squamous Cell Carcinoma	22 (16.3%)
Adenoid Cell Carcinoma	48 (35.5%)
Large Cell Carcinoma	7 (5.2%)
Mixture of Squamous and Adenoid Cell Carcinoma	2 (1.5%)
Diffuse Large B Lymphoma	1 (0.7%)
Thymoma	1 (0.7%)
Gender	**Number (Percentage)**
Male	87 (64%)
Female	48 (36%)
Age (years, mean ± standard deviation)	58 ± 11
Age Range (years)	17-85

Note: granulomatous lesion, inflammatory pseudotumor, hemangioma and fibrocellular tumor are benign lung lesions, while squamous cell carcinoma, adenoid cell carcinoma, large cell carcinoma, mixture of squamous and adenoid cell carcinoma, diffuse large B lymphoma and thymoma are malignant lung lesions.

Table 2
Five groups of CT-radiomics features selected by the least absolute shrinkage and selection operator (LASSO) and Forward selection approach.

Group	CT-radiomics features
group 0	Original_shape_SurfaceVolumeRatio
	Logarithm_firstorder_RootMeanSquared
group 1	Original_shape_SurfaceVolumeRatio
	Original_shape_Sphericity
	Logarithm_glcm_ClusterShade
	Wavelet_LLH_glrlm_ShortLowGrayLevelEmphasis
group 2	Original_shape_SurfaceVolumeRatio
	Logarithm_firstorder_RootMeanSquared
	Original_shape_Elongation
group 3	Original_shape_SurfaceVolumeRatio
	Logarithm_firstorder_RootMeanSquared
	Wavelet_HHL_firstorder_Maximum
	Wavelet_HHH_glszm_HighGrayLevelZoneEmphasis
	Wavele_HHL_glszm_SmallAreaHighGrayLevelEmphasis
group 4	Original_shape_SurfaceVolumeRatio
	Original_shape_Compactness1

Note: Group, from group 0 to group 4, means the training set in each run of five-fold cross-validation. The meaning of each feature format is Filter_FeatureGroup_FeatureName, and 'Original' stands for the radiomics features extracted from the original images without preprocessing.

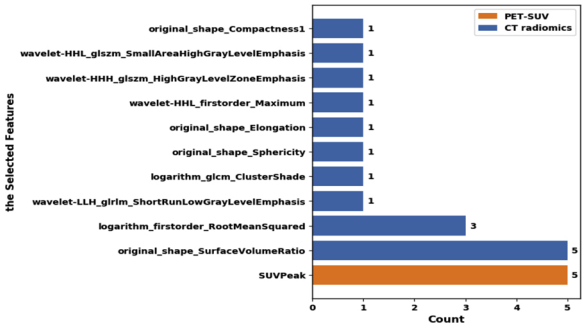

Fig. 2. The bar chart of counts of all selected CT-radiomics features and PET metabolic parameters in five groups. 'Orange' bar depicts the only selected metabolic parameter – SUVpeak; 'Blue' bars depict the radiomics features selected in five groups, and ten features appear at least once with 'logarithm_firstorder_RootMeanSquared' and 'original_shape_SurfaceVolumnRatio' being more frequent.

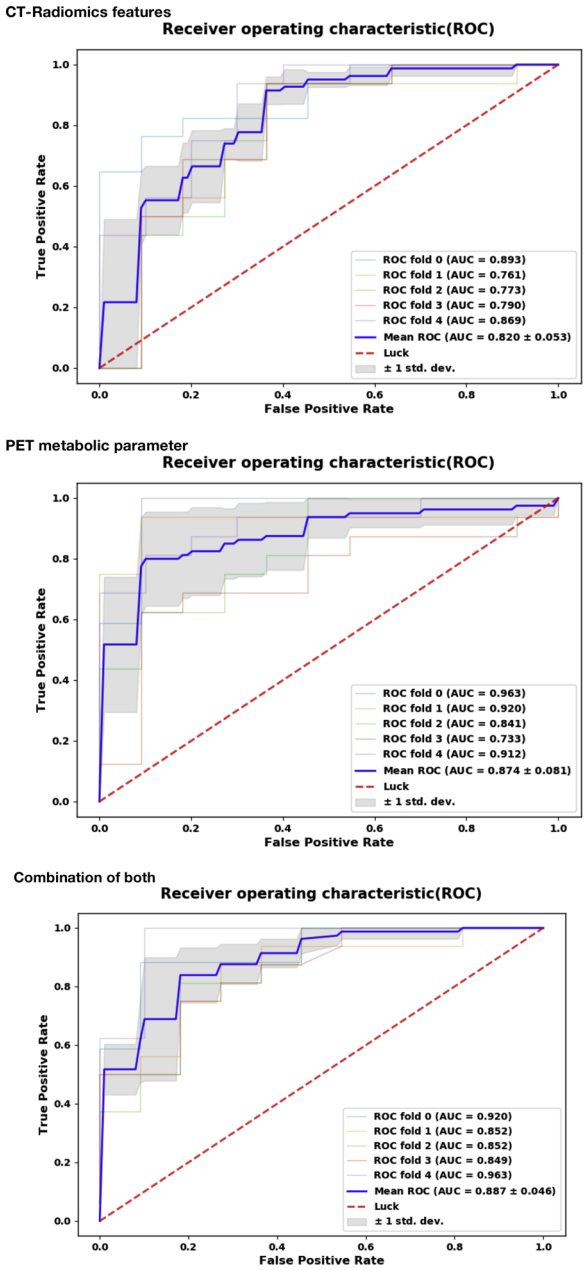

Fig. 3. Receiver operating characteristic (ROC) curves for identifying benign and malignant lung lesions using CT-radiomics features, PET metabolic parameters and the combination of both models, respectively.
In each subplot, five light thin curves respectively represent the ROC of each fold (namely, test set) in five-fold cross-validation, and the blue thick curve is mean ROC curve with the shadow area of standard deviation. The dotted line named 'Luck' showed that inputted features could not distinguish two groups and the area is equal to 0.5. More generally, whether the prediction was benign or malignant is by a matter of 'Luck'.

3.3. Diagnostic performance

The diagnostic performance of three kinds of models including radiomics parameters, PET metabolism parameter and the combination of both illuminated by ROC curves is shown in Fig. 3, with mean AUC of 0.820 ± 0.053 and 0.874 ± 0.081 and 0.887 ± 0.046, respectively. Mean AUC, accuracy and sensitivity and specificity of five-fold cross-validation are summarized in Table 3.

In each run, no significant difference among ROC curves of CT-radiomics features, PET metabolic parameters and combined models (range of P-value from 0.09 to 0.81, Delong test) was observed. When comparing the ROC curves of combining the predicted probabilities of all cases, we found significant differences between those obtained from CT-radiomics features and those obtained from combined models (P = 0.018, Delong test).

3.4. Comparison of benign and malignant groups based on potential features

Differentiation of benign and malignant lung lesions based on both 'SurfaceVolumeRatio' (P < 0.001, U test) and 'SUVpeak' (P < 0.001, U test) showed significant differences (shown in Fig. 4). The histograms and distribution density of potential features between benign and malignant groups are shown in Fig. 5**A** and **B**. For 'SUVpeak', the distribution density of two groups was estimated by a Gaussian kernel with bandwidth of 0.04, the values of benign mainly concentrated in the range 0–10, while the values of malignant tended to be larger ranging from 5 to 15.

The distribution density of 'SurfaceVolumeRatio' was fitted by same kernel with bandwidth of 1.8, and the value range of malignant was 0.1-0.4 approximately. We noticed that the distribution density of 'SurfaceVolumeRatio' for the benign group had two peaks, and the benign lung lesions mainly contained two kinds of subtypes named "granulomatous lesion" and "inflammatory pseudotumor". Consequently, the U test and distribution density estimation of potential features were also conducted in the granulomatous lesion and inflammatory pseudotumor subtypes, and the consequences showed a significant difference (SurfaceVolumeRatio, P = 0.00029; SUVpeak, P = 0.030, U test; shown in Fig. 4) and one peak on each of the two subtypes (shown in Fig. 5C).

4. Discussion

The study developed three kinds of classification models using an SVM method and the strategy of five-fold cross-validation to distinguish the properties of lung lesions. These models were the models of CT-radiomics features, PET metabolic parameters and the combination of both respectively. The results showed satisfactory diagnostic performances for both CT-radiomics features and PET metabolic parameters with mean AUC of 0.820 ± 0.053, 0.874 ± 0.081 and 0.887 ± 0.046 for the combination of both.

Despite of this, there was the significant difference between CT-radiomics features and the combination on the whole data set (P = 0.018, Delong test). We did not observe any significant differences in ROC curves among the models through pairwise comparison in each fold (range of P-value from 0.09 to 0.81, Delong test). Considering the intuitive difference in AUC values, we generally think that the performance from high to low is the combination model, model followed by PET metabolic parameters and then CT-radiomics features, which was almost consistent with the report of Shanna et al. [7], if we must sort the three models. Shanna et al. [7] reported that PET/CT performs significantly better than PET alone in accuracy, sensitivity and specificity, and better than CT alone in specificity. The integration of advantage with PET/CT is better selection by maintaining the sensitivity of CT and the specificity of PET. In our study, radiomics features was used to obtain more information of lesions on PET/CT images.

In clinical practice, it is, of course, ideal to use the both modalities of PET and CT modalities or a single PET for pulmonary lesion diagnosis, but in view of the high penetration rate and low cost of CT around the world, our study shows that CT based on radiomics features might be a good alternative under the limited medical conditions.

Previous studies [14,30,31] reported that radiomics in medical imaging had been successfully applied to non-invasive diagnosis and prognosis, including the prediction of pulmonary lesions [13,14,31–33]. In line with those similar researches, we almost used the followed methods: 1) CT-radiomics to extract a variety of image features including texture, intensity and shape; 2) classical and efficient

Table 3
Diagnostic performance of CT-radiomics features, PET metabolic parameters and the combined of both models in five-fold cross-validation.

Model	AUC	Accuracy	Sensitivity	Specificity
CT-Radiomics	0.820 ± 0.053	0.740 ± 0.043	0.740 ± 0.050	0.740 ± 0.040
PET metabolic parameter	0.874 ± 0.081	0.814 ± 0.085	0.814 ± 0.089	0.815 ± 0.082
Combined	0.887 ± 0.046	0.815 ± 0.066	0.814 ± 0.058	0.816 ± 0.079

Note: the format of value was mean ± standard deviation, AUC is area under receiver operating characteristic curve.

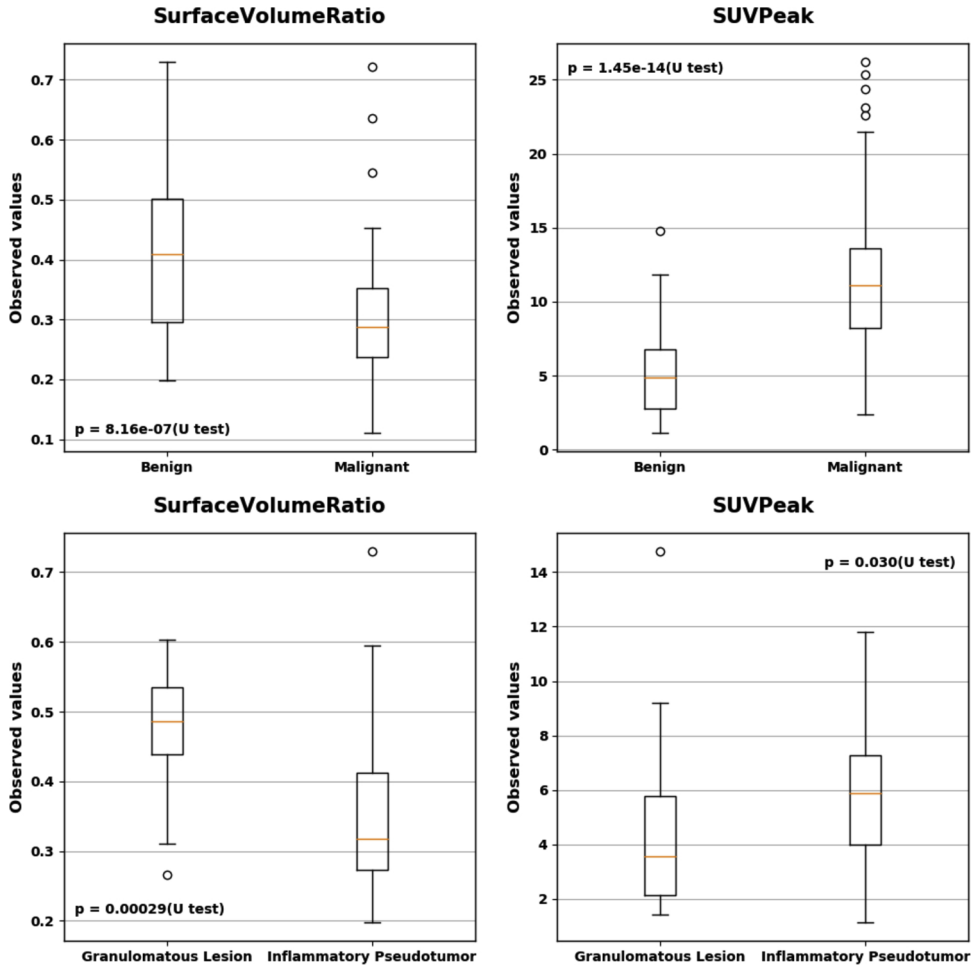

Fig. 4. The box plots of potential features (SurfaceVolumeRatio in the first column and SUVpeak in the second column) between benign and malignant groups, granulomatous lesion and inflammatory pseudotumor subtypes. It showed significant differences between benign and malignant groups, and granulomatous lesion and inflammatory pseudotumor in SurfaceVolumeRatio and SUVpeak using the Mann-Whitney U test. The hollow circles represent the outliers beyond normal ranges in their datasets.

machine learning methods for feature selection (LASSO in Choi W et al. [13] and us) and model building (such as SVM in Choi W et al. and Hawkins S et al. [13,14] and us, random forest(RF) in Huang P et al. [31] and Hawkins S et al. [14]); 3) using nested cross-validation methods to maximize the utility of limited data sets while ensuring repeatability and stability of results (except Huang P et al. [31], which just split the whole data set into training set and validation set).

However, there remain some striking differences between previous studies and ours. Firstly, the goal of reports of Huang P et al. [31] Choi W et al. [13] and Hawkins S et al. [14] was to predict and detect early the potential malignant pulmonary nodules (namely, lung cancer), resulting in the small nodule lengths. Consequently, they used relatively short ranging from 4 to 20 mm, while our study aimed to determine the properties of lung lesions and included the lung lesions with volume more than 1cm^3. Secondly, the studies [32,33] in the past preferred using PET images for predicting tumor histopathology and genetics instead of classifying lung lesions, which might due to the limited resolution of PET for detection of small nodules and rare studies focusing on the determination of properties of lesions. But we used PET/CT images to expand research materials and ensure the consistency of data and the accuracy of lesion segmentation.

We also explored the potential features for diagnosis of lesion properties determined by cross-validation, which were 'SurfaceVolumeRatio' from CT-radiomics features belonging to the group of shape, and 'SUVpeak' from PET metabolic parameters. The 'SurfaceVolumeRatio' stands for the ratio between surface and volume for an entity, for which the larger the value is, the more blurred and lower the edges is. The finding in our study was consistent with those in these reports [32,34] using fractal analysis. Significant differences between two groups (benign and malignant lung lesions) and potential features were also found (SUVpeak: $P < 0.001$, SurfaceVolumeRatio: $P < 0.001$, U test), thus further supporting the hypothesis that the

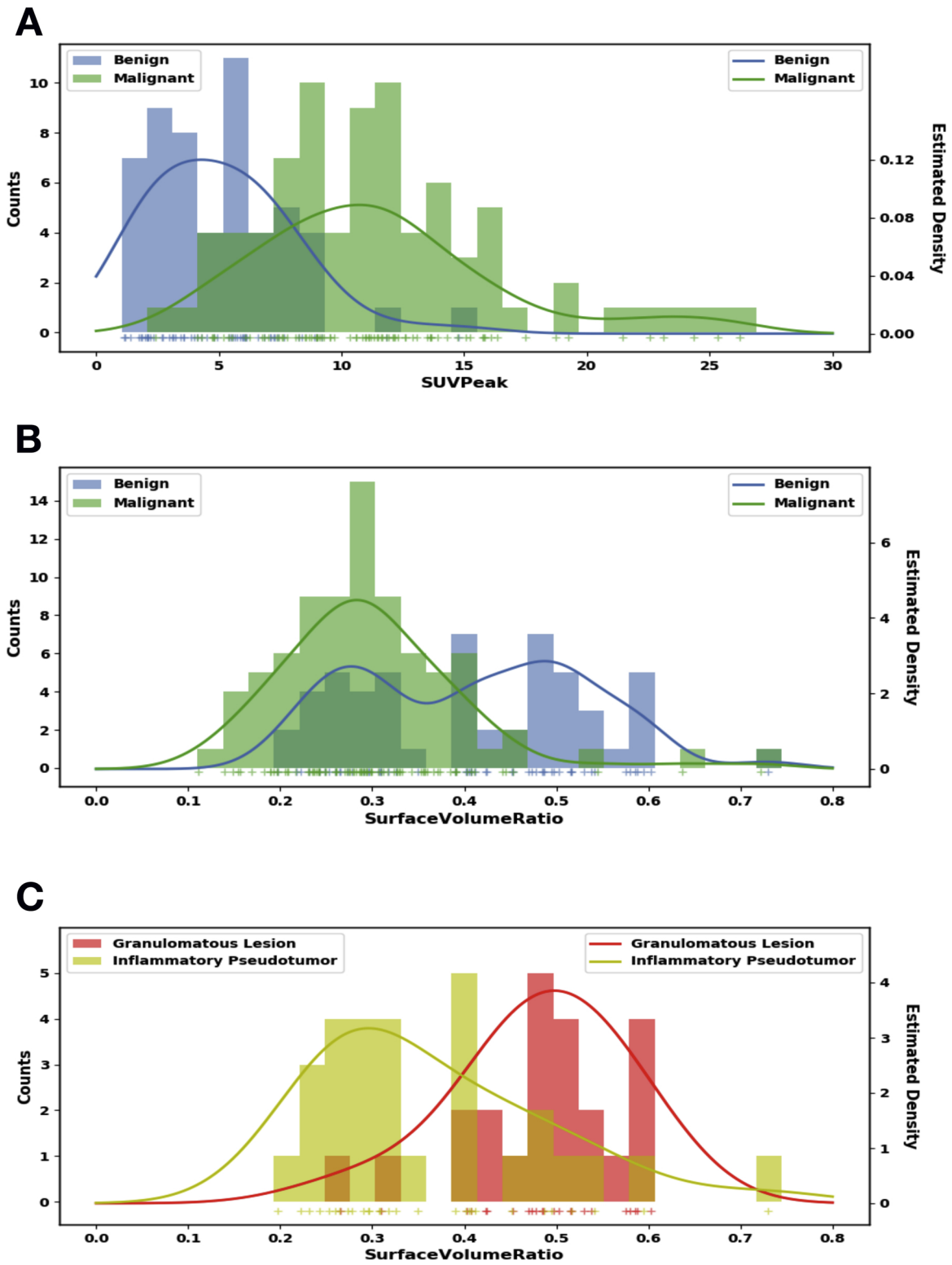

Fig. 5. The histograms and estimated distribution density of potential features (**A**, SUVpeak; **B**, SurfaceVolumeRatio) between benign and malignant groups, and (**C**) SurfaceVolumeRatio between granulomatous lesion and inflammatory pseudotumor subtypes. In each subplot, dual-Y axis was used to indicate two forms, in which the left one stands for the counts of each bar in histogram and the right one means the density of distribution estimated by Gaussian kernel with appropriate bandwidth of 0.04 (**A**), 1.8 (**B**) and 1.8 (**C**).

combination of CT-radiomics features and PET metabolic parameters could be useful as diagnostic tool. Furthermore, when comparing the potential CT-radiomics features selected by different studies, we noticed that there were more selected textural features and different between each other in published reports [13,14,31]; instead our study just acquired one most valuable radiomics feature which belonged to shape feature. According to our speculation, the reasons might be that 'SurfaceVolumeRatio' is not significant enough in the small length nodules, the retrospective study cannot completely control various the interference factors and the texture feature is greatly unstable for the

disturbance fluctuation [35,36].

For explanation of the distribution density of potential features for benign lesions with two peaks, further analysis was conducted to distinguish granulomatous lesion from inflammatory pseudotumor using U test and distribution density estimation. The results showed a statistically significant difference for 'SurfaceVolumeRatio' ($P = 0.00029$) and 'SUVpeak' ($P = 0.030$), but poorer capacity for distinguishing between granulomatous lesions from inflammatory pseudotumor using PET parameters than CT-radiomics features. A possible explanation might be that we only included patients with primary tumor of volume of at least $1cm^3$. In accordance with previous publications, the minimum size was chosen in order to reliably quantify tracer uptake heterogeneity parameters [37], this patient cohort may also have had an inflammatory component. Although encouraging, the potential features acquired were possible following data augmentation, which could potentially lead to overoptimistic results.

This study was a retrospective analysis that enrolled a small number of diverse patients with different pathologies of benign or malignant lesions. Even though the large number of 18 F-FDG PET/CT images are available in our institute, the variety of parameters of equipment setup exists not only among different facilities but also within the same institution. Studies reported the great issue on reliability and repeatability of radiomics extraction parameters from CT images [38,39], and the effect of imaging parameters including scanning parameters and section thickness [40]. The efforts we made were to eliminate the uncertainty and instability of the features as much as possible, such as using the nested cross-validation instead of separating another validation set for the limited data, image preprocessing and building segmentation standard for extracting lung lesion features on images. In order to utilize for clinical care widely, our future studies should focus on further standardization of image acquisition, reconstruction setting, segmentation methods and radiomics analysis [41,42]. And data balance in different pathology of lung lesion need be further considered [43,44].

5. Conclusion

The diagnostic performances of CT-radiomics features, PET metabolic parameters and combination models developed and analyzed based on 18 F-FDG PET/CT images were greatly satisfactory. Use of the SurfaceVolumeRatio from the CT-radiomics features and SUVpeak from the PET metabolic parameters have substantial capacity for differentiation between benign and malignant lung lesions. Moreover, SurfaceVolumeRatio has distinctive characterization to differentiate granulomatous lesion from inflammatory pseudotumor.

Ethical approval and consent to participate

The study was approved by the Tianjin Cancer Hospital Medical Ethics Committee and informed consent was waived.

Funding

The work is supported by the Joint Large-Scale Scientific Facility Funds of the NSFC and CAS [U1532258].

Declaration of Competing Interest

All the authors listed above had fully participated in the study, read and approved the final version of the manuscript and consented for publication. We declared there exist no possible conflicts of interest.

Acknowledgements

We thank Dr. George Starkschall from Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX77030, for his help with the revision of manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2019.108735.

References

[1] N. Ezer, A. Navasakulpong, K. Schwartzman, L. Ofiara, A.V. Gonzalez, Impact of rapid investigation clinic on timeliness of lung cancer diagnosis and treatment, BMC Pulm. Med. 17 (December (1)) (2017) 178.
[2] D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, R.M. Fagerstrom, I.F. Gareen, C. Gatsonis, P.M. Marcus, J.R.D. Sicks, Reduced lung-cancer mortality with low-dose computed tomographic screening, N. Engl. J. Med. 365 (August (5)) (2011) 395–409.
[3] S. Tan, S. Kligerman, W. Chen, M. Lu, G. Kim, S. Feigenberg, W.D. D'Souza, M. Suntharalingam, W. Lu, Spatial-temporal [18F]FDG-PET features for predicting pathologic response of esophageal Cancer to neoadjuvant chemoradiation therapy, Int. J. Radiat. Oncol. Biol. 85 (April (5)) (2013) 1375–1382.
[4] M.L. Rosado de Christenson, A comparison of the diagnostic accuracy of 18F-FDG PET and CT in the characterization of solitary pulmonary nodules, Yearbook of Diagnostic Radiology vol. 2009, (2009), pp. 11–13 Jan..
[5] F. Tixier, M. Hatt, C. Valla, V. Fleury, C. Lamour, S. Ezzouhri, P. Ingrand, R. Perdrisot, D. Visvikis, C.C. Le Rest, Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: prognostic value in non-small cell lung Cancer, J. Nucl. Med. 55 (August (8)) (2014) 1235–1241.
[6] O. van Gómez López, A.M.G. Vicente, A.F.H. Martínez, Á.M.S. Castrejón, G.A.J. Londoño, J.M. Udias, P.L. Atance, Heterogeneity in [18F] Fluorodeoxyglucose positron emission tomography/computed tomography of non–small cell lung carcinoma and its relationship to metabolic parameters and pathologic staging, Mol. Imaging 13 (November (9)) (2014) 7290 00032. 2014.
[7] S.K. Kim, M. Allen-Auerbach, J. Goldin, B.J. Fueger, M. Dahlbom, M. Brown, J. Czernin, C. Schiepers, Accuracy of PET/CT in characterization of solitary pulmonary lesions, J. Nucl. Med. 48 (February (2)) (2007) 214–220.
[8] Z. Ma, M. Fang, Y. Huang, L. He, X. Chen, C. Liang, X. Huang, Z. Cheng, D. Dong, C. Liang, J. Xie, J. Tian, Z. Liu, CT-based radiomics signature for differentiating Borrmann type IV gastric cancer from primary gastric lymphoma, Eur. J. Radiol. 91 (June) (2017) 142–147.
[9] X. Gao, C. Chu, Y. Li, P. Lu, W. Wang, W. Liu, L. Yu, The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from (18)F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer, Eur. J. Radiol. 84 (February (2)) (2015) 312–317.
[10] M.B. Andersen, S.W. Harders, B. Ganeshan, J. Thygesen, H.H. Torp Madsen, F. Rasmussen, CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer, Acta Radiol. 57 (June (6)) (2016) 669–676.
[11] H. Bayanati, R.E. Thornhill, C.A. Souza, V. Sethi-Virmani, A. Gupta, D. Maziak, K. Amjadi, C. Dennie, Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? Eur. Radiol. 25 (Feb (2)) (2015) 480–487.
[12] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (February (2)) (2016) 563–577.
[13] W. Choi, J.H. Oh, S. Riyahi, C.-J. Liu, F. Jiang, W. Chen, C. White, A. Rimner, J.G. Mechalakos, J.O. Deasy, W. Lu, Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer, Med. Phys. 45 (April (4)) (2018) 1537–1549.
[14] S. Hawkins, H. Wang, Y. Liu, A. Garcia, O. Stringfield, H. Krewer, Q. Li, D. Cherezov, R.A. Gatenby, Y. Balagurunathan, D. Goldgof, M.B. Schabath, L. Hall, R.J. Gillies, Predicting malignant nodules from screening CT scans, J. Thorac. Oncol. 11 (December (12)) (2016) 2120–2128.
[15] S.R. Digumarthy, A.M. Padole, R. Lo Gullo, R. Singh, J.-A.O. Shepard, M.K. Kalra, CT texture analysis of histologically proven benign and malignant lung lesions, Medicine (Baltimore) 97 (June (26)) (2018) e11172.
[16] A. Oikonomou, F. Khalvati, P.N. Tyrrell, M.A. Haider, U. Tarique, L. Jimenez-Juan, M.C. Tjong, I. Poon, A. Eilaghi, L. Ehrlich, P. Cheung, Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy, Sci. Rep. 8 (March (1)) (2018) 4003.
[17] T. Merlin, D. Visvikis, P. Fernandez, F. Lamare, A novel partial volume effects correction technique integrating deconvolution associated with denoising within an iterative PET image reconstruction, Med. Phys. 42 (February (2)) (2015) 804–819.
[18] Y. Tan, L.H. Schwartz, B. Zhao, Segmentation of lung lesions on CT scans using watershed, active contours, and Markov random field, Med. Phys. 40 (April (4)) (2013) 043502.
[19] J.J.M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, H.J.W.L. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (November (21)) (2017) e104–e107.
[20] Z. Yaniv, B.C. Lowekamp, H.J. Johnson, R. Beare, SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research, J. Digit. Imaging 31 (June (3)) (2018) 290–303.
[21] M. Dash, H. Liu, Feature selection for classification, Intell. Data Anal. 1 (1) (1997) 131–156.
[22] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (January (3)) (1995) 273–297.

[23] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. 19 (December (6)) (1974) 716–723.
[24] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (1) (2010) 1–22.
[25] J. Kittler, P.A. Devijver, Statistical properties of error estimators in performance assessment of recognition systems, IEEE Trans. Pattern Anal. Mach. Intell. 4 (2) (1982) 215–220.
[26] C. Campbell, Y. Ying, Learning with support vector machines, Learn. Supp. Vector Mach. 5 (1) (2011) 1–95.
[27] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (January (1)) (1950) 32–35.
[28] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (September (3)) (1988) 837.
[29] H.B. Mann, D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18 (March (1)) (1947) 50–60.
[30] R. Paul, S.H. Hawkins, M.B. Schabath, R.J. Gillies, L.O. Hall, D.B. Goldgof, Predicting malignant nodules by fusing deep features with classical radiomics features, J. Med. Imaging 5 (1) (2018) 1.
[31] P. Huang, S. Park, R. Yan, J. Lee, L.C. Chu, C.T. Lin, A. Hussien, J. Rathmell, B. Thomas, C. Chen, R. Hales, D.S. Ettinger, M. Brock, P. Hu, E.K. Fishman, E. Gabrielson, S. Lam, Added value of computer-aided CT image features for early lung Cancer diagnosis with small pulmonary nodules: a matched case-control study, Radiology 286 (January (1)) (2018) 286–295.
[32] U. Bashir, M.M. Siddique, E. Mclean, V. Goh, G.J. Cook, Imaging heterogeneity in lung Cancer: techniques, applications, and challenges, Am. J. Roentgenol. 207 (September (3)) (2016) 534–543.
[33] C. Hassani, B.A. Varghese, J. Nieva, V. Duddalwar, Radiomics in pulmonary lesion imaging, Am. J. Roentgenol. 212 (March (3)) (2019) 497–504.
[34] S. Kido, K. Kuriyama, M. Higashiyama, T. Kasugai, C. Kuroda, Fractal analysis of internal and peripheral textures of small peripheral bronchogenic carcinomas in thin-section computed tomography: comparison of bronchioloalveolar cell carcinomas with nonbronchioloalveolar cell carcinomas, J. Comput. Assist. Tomogr. 27 (January (1)) (2003) 56–61.
[35] R. Berenguer, M.D.R. Pastor-Juan, J. Canales-Vázquez, M. Castro-García, M.V. Villas, F. Mansilla Legorburo, S. Sabater, Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters, Radiology 288 (August (2)) (2018) 407–415.
[36] X. Fave, D. Mackin, J. Yang, J. Zhang, D. Fried, P. Balter, D. Followill, D. Gomez, A.K. Jones, F. Stingo, J. Fontenot, L. Court, Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? Med. Phys. 42 (December (12)) (2015) 6784–6797.
[37] M. Hatt, M. Majdoub, M. Vallieres, F. Tixier, C.C. Le Rest, D. Groheux, E. Hindie, A. Martineau, O. Pradier, R. Hustinx, R. Perdrisot, R. Guillevin, I. El Naqa, D. Visvikis, 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort, J. Nucl. Med. 56 (1) (2014) 38–44.
[38] F.H.P. van Velden, G.M. Kramer, V. Frings, I.A. Nissen, E.R. Mulder, A.J. de Langen, O.S. Hoekstra, E.F. Smit, R. Boellaard, Repeatability of radiomic features in non-small-cell lung cancer [(18)F]FDG-PET/CT studies: impact of reconstruction and delineation, Mol. Imaging Biol. 18 (October (5)) (2016) 788–795.
[39] D. Mackin, X. Fave, L. Zhang, D. Fried, J. Yang, B. Taylor, E. Rodriguez-Rivera, C. Dodge, A.K. Jones, L. Court, Measuring computed tomography scanner variability of radiomics features, Invest. Radiol. 50 (November (11)) (2015) 757–765.
[40] L. He, Y. Huang, Z. Ma, C. Liang, C. Liang, Z. Liu, Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule, Sci. Rep. 6 (October (1)) (2016) 161.
[41] M. Hatt, F. Tixier, L. Pierce, P.E. Kinahan, C.C. Le Rest, D. Visvikis, Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur. J. Nucl. Med. Mol. Imaging 44 (January (1)) (2017) 151–165.
[42] R. Boellaard, R. Delgado-Bolton, W.J.G. Oyen, F. Giammarile, K. Tatsch, W. Eschner, F.J. Verzijlbergen, S.F. Barrington, L.C. Pike, W.A. Weber, S. Stroobants, D. Delbeke, K.J. Donohoe, S. Holbrook, M.M. Graham, G. Testanera, O.S. Hoekstra, J. Zijlstra, E. Visser, C.J. Hoekstra, J. Pruim, A. Willemsen, B. Arends, J. Kotzerke, A. Bockisch, T. Beyer, A. Chiti, B.J. Krause, European Association of Nuclear Medicine (EANM), "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0, Eur. J. Nucl. Med. Mol. Imaging 42 (February (2)) (2015) 328–354 Springer Berlin Heidelberg.
[43] D. Fehr, H. Veeraraghavan, A. Wibmer, T. Gondo, K. Matsumoto, H.A. Vargas, E. Sala, H. Hricak, J.O. Deasy, Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images, Proc. Natl. Acad. Sci. U. S. A. 112 (November (46)) (2015) E6265–73.
[44] Y. Zhang, A. Oikonomou, A. Wong, M.A. Haider, F. Khalvati, Radiomics-based prognosis analysis for non-small cell lung cancer, Sci. Rep. 7 (April (1)) (2017) 46349.