Development and Validation of a 18F-FDG PET-Based Radiomic Model for Evaluating Hypermetabolic Mediastinal–Hilar Lymph Nodes in Non-Small-Cell Lung Cancer

Ming-li Ouyang1, Yi-ran Wang2, Qing-shan Deng3, Ye-fei Zhu4, Zhen-hua Zhao5, Ling Wang6, Liang-xing Wang1* and Kun Tang6*

1 Department of Respiratory Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, 2 Department of Medical Engineering, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, 3 Department of Radiology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, 4 Department of Respiratory Medicine, Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Taizhou, China, 5 Department of Radiology, Shaoxing People’s Hospital, Shaoxing Hospital of Zhejiang University, Shaoxing, China, 6 Department of Nuclear Medicine, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

Background: Accurate evaluation of lymph node (LN) status is critical for determining the treatment options in patients with non-small cell lung cancer (NSCLC). This study aimed to develop and validate a 18F-FDG PET-based radiomic model for the identification of metastatic LNs from the hypermetabolic mediastinal–hilar LNs in NSCLC.

Methods: We retrospectively reviewed 259 patients with hypermetabolic LNs who underwent pretreatment 18F-FDG PET/CT and were pathologically confirmed as NSCLC from two centers. Two hundred twenty-eight LNs were allocated to a training cohort (LN = 159) and an internal validation cohort (LN = 69) from one center (7:3 ratio), and 60 LNs were enrolled to an external validation cohort from the other. Radiomic features were extracted from LNs of PET images. A PET radiomics signature was constructed by multivariable logistic regression after using the least absolute shrinkage and selection operator (LASSO) method with 10-fold cross-validation. The PET radiomics signature (model 1) and independent predictors from CT image features and clinical data (model 2) were incorporated into a combined model (model 3). A nomogram was plotted for the complex model, and the performance of the nomogram was assessed by its discrimination, calibration, and clinical usefulness.

Results: The area under the curve (AUC) values of model 1 were 0.820, 0.785, and 0.808 in the training, internal, and external validation cohorts, respectively, showing good diagnostic efficacy for lymph node metastasis (LNM). Furthermore, model 2 was able to discriminate metastatic LNs in the training (AUC 0.780), internal (AUC 0.794), and external validation cohorts (AUC 0.802), respectively. Model 3 showed optimal diagnostic performance among
INTRODUCTION

Lung cancer is still the leading cause of cancer-related mortality worldwide (1). Non-small-cell lung cancer (NSCLC) accounts for about 85% of lung cancers (2). The occurrence of contralateral or multiregional mediastinal–hilar lymph node metastasis (LNM) in NSCLC might exclude the patient from primary surgery (3, 4), which is significantly associated with unfavorable clinical prognosis. Therefore, accurate lymph nodal staging is critical for determining the treatment options in patients with NSCLC for clinicians.

In recent years, 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography/computed tomography (PET/CT) has played an important role in evaluating lymph node (LN) status, which can provide both anatomic and glucose metabolic information (5). Numerous studies have shown that intrathoracic nodal status is considered to be positive for metastatic spread if the activity of the node was higher than the mediastinal background (6–8). However, a high level of FDG uptake can also be detected in benign LNs, illustrating the risk of misjudging inflammatory and reactive lymph nodes for metastatic lymph nodes (9, 10). The variable positive predictive rate for assessment of LN involvement with PET/CT has also been recognized, ranging from 32.3% to 89% (9–11). Obviously, the false-positive results of patients with NSCLC could misguide clinicians in making treatment decisions, such as missing surgery or more aggressive treatments, and ultimately influence prognosis. Although PET/CT has a good negative predictive value for the investigation of LNM (9), we have done some studies and published several papers (12, 13). Therefore, various practice guidelines recommend invasive staging procedures with potential side effects for patients with PET positive lymph nodes (14).

To exhaust the full potential information of noninvasive radiology images, a more sophisticated approach is urgently needed. Radiomics is a recently emerging technique for high-throughput mining of tumor characteristics from the images, which reflects the underlying tumor biology and behavior (15, 16). In the field of radiomics analysis based on PET/CT, there has been recently increasing evidence demonstrating the advantages of radiomics as a noninvasive approach in the diagnosis, staging, prognosis, and treatment outcome of lung cancer (13, 17–19). Simultaneously, several studies have focused on assessing the lymph node metastasis using the machine learning algorithm in NSCLC. Yin et al. (20) built a support vector machine model to predict the mediastinal LNM.

the three cohorts, with an AUC of 0.874, 0.845, and 0.841, respectively. The nomogram based on the model 3 showed good discrimination and calibration.

Conclusions: Our study revealed that PET radiomics signature, especially when integrated with CT imaging features, showed the ability to identify true and false positives of mediastinal–hilar LNM detected by PET/CT in patients with NSCLC, which would help clinicians to make individual treatment decisions.

Keywords: non-small cell lung cancer (NSCLC), hypermetabolic lymph node, metastasis, positron emission tomography/computed tomography (PET/CT), radiomics

MATERIALS AND METHODS

Patient and Lymph Node Selection

In this retrospective study, we screened and collected 18F-FDG PET/CT and clinical data about patients as the primary cohort in The First Afﬁliated Hospital of Wenzhou Medical University from January 2012 to September 2020. The patients for the external validation cohort were selected from Shaoxing People’s Hospital from November 2016 to December 2020. The study was approved by The Institutional Review Boards of the two participating institutions, and informed consent was waived for this retrospective study.

The specific inclusion criteria were as follows: 1) with pathological diagnosis of NSCLC, 2) with systematic hilar and mediastinal lymph node dissection or the positive results of endobronchial ultrasound-guided transbronchial needle aspirate (EBUS-TBNA) biopsy due to the high false-negative rates [Several studies showed that the positive predictive value and negative predictive values of EBUS-TBNA ranged 92%–100% and 11–97%, respectively (22, 23)], and 3) underwent 18F-FDG PET/CT examination. The exclusion criteria for patients were 1) the interval between the pathologic findings and PET/CT scan more than 3 weeks, 2) the patient without the increased FDG uptake LNs, 3) with distant metastasis, 4) with history of other cancer, 5) received neoadjuvant chemoradiotherapy, and 6) images with poor quality as follows: 1) the leakage of 18F-FDG at the injection site, which leads to a decrease of FDG uptake of images; 2) images with inferior quality due to noise, respiratory artifacts, or other movement artifacts; and 3) missing PET images or CT images. The exclusion criteria for LNs were 1) mediastinal–hilar LN uptake below or comparable to mediastinal background activity, 2) the boundary...
between the hypermetabolic LNs and the primary lesions not clear on PET images, 3) with a short-axis diameter of LNs less than 5 mm insufficient to outline a valid volume of interest (VOI), and 4) VOIs did not have enough voxels (at least 64 voxels).

Specific process of determining the pathology of LNs was as follows: 1) radiological, surgical, and pathological LN division all followed the eighth edition of the Union for International Cancer Control TNM classification. The histological findings served as the reference standard for comparison with the PET/CT findings (2). When the LNs dissected in the operation or performed by EBUS-TBNA were pathologically proven to be metastatic in one region, then the identified LNs with hypermetabolic activity on PET images in the same region were regarded as metastatic. 3) If all the harvested LNs within surgery were pathologically proven to be non-metastatic in one region, then the identified hypermetabolic LNs on PET images in the same region were considered as non-metastatic. 4) If systematic LN dissection did not include one region, then the hypermetabolic LNs on PET images in this region were excluded. The detailed flowchart of enrollment is shown in Supplementary Material Image 1.

According to the above criteria, 228 lymph nodes (219 NSCLC patients) were identified as the primary cohort, which were assigned to a training cohort (LN = 159) and an internal validation cohort (LN = 69) by the random split-sample (7:3) method (Figure 1). Sixty lymph nodes (40 NSCLC patients) were included in the external validation cohort.

**PET/CT Protocol**

18F-FDG PET/CT scanning was performed on either Gemini TF 64 (Philips, the Netherlands, 219 patients, 84.6%) from The First Affiliated Hospital of Wenzhou Medical University or General Electric Discovery Elite (Waukesha, WI, 40 patients, 15.4%) from Shaoxing People’s Hospital. All patients were instructed to fast for at least 6 h (blood glucose levels below 110 ml/dl) before being injected with 18F-FDG (3.7–5.55 MBq/kg). PET/CT images were acquired from the skull base to the mid-thigh after the injection with an average interval of 59.6 (± 7.8) min. Attenuation correction of the PET scan was based on the CT images. The detailed acquisition parameters are given in Supplementary Material Table 1, according to the Image Biomarker Standardization Initiative (IBSI) guidelines (24).

**PET Radiomic Feature Extraction**

FDG-PET images with the DICOM format were imported into LIFEx freeware (version 6.30, http://www.lifexsoft.org), and the software was used to extract radiomics features. The VOI of the LN was manually delineated and segmented slice by slice using 3D drawing tools by two experienced nuclear medicine physicians who were blinded to pathological results. Forty percent of the maximum standardized uptake value (SUVmax) was adopted as a threshold to optimize the VOI. VOIs were placed to avoid adjacent vessels, bronchi, and the chest wall. For image preprocessing, the resampling voxel size was set at 4 × 4 × 4 mm on PET images in all the patients. We chose an isotropic voxel size of 4 mm because the original voxel size was closer to 4 mm in the two centers of our study. Intensity discretization and rescaling were performed automatically by the software. For PET images, intensity discretization was processed by decreasing the continuous scale to 64 bins, and the intensity rescaling bonds ranged from 0 to 20. Radiomic features were divided into seven groups: conventional
features (e.g., SUVmax, SUVmean), discretized/histogram (HISTO), shape, gray-level zone-length matrix (GLZLM), gray-level run-length matrix (GLRLM), neighborhood gray-level different matrix (NGLDM), and gray-level co-occurrence matrix (GLCM). Radiomic features extracted from PET images are shown in Supplementary Material Table 2. Source data are available in Supplementary Material Table 3.

CT Image Analysis
Based on the CT component of the PET/CT images, the analysis of the image features was performed by two experienced nuclear medicine physicians. In case of disagreements, consensus was sought by discussion. The following conventional CT features about LN were analyzed: size, shape, margin, calcification, cystic change, CTmax, and CTmean. The size of the lymph node was defined by the maximal short-axis diameter. We defined round or oval as regular shape according to CT images, and any other shape was defined as irregular. CTmax and CTmean were measured at the largest section of LN on transverse CT images.

Radiomic Feature Screening and Signature Building
To verify interobserver reliability, the intraclass correlation coefficient (ICC) was performed for each VOI-based radiomic feature. Features with ICC greater than 0.8 were selected for subsequent investigation. Additionally, least absolute shrinkage and selection operator (LASSO) regression was applied to further screen the optimal subset from the training cohort and develop a formula by multivariable logistic regression to calculate the radiomics score (Rad-score). Ten-fold cross-validation was used to tune the optimal lambda (λ) value that was the minimum mean cross-validated error.

Construction of Three Prediction Models

Model 1: PET Radiomics Signature
In the training cohort, the optimal PET radiomics features identified by the above LASSO logistic regression procedure using 10-fold cross-validation were utilized to construct a radiomics signature (Rad-score), which formed the prediction model 1.

Model 2: CT Image Features and Clinical Data
The differences between LNM and non-LNM groups in CT image features and clinical data were firstly assessed using univariable logistic regression analysis. Features with $p < 0.05$ in the univariate analysis were included in subsequent multivariable analysis, and model 2 was built in the training cohort using the backward step-down process.

Model 3: Combination of Model 1 and Model 2
The complex model (model 3) was the combination of model 1 and model 2 by multivariable logistic regression using backward stepwise selection in the training cohort.

Validation of the Models and Clinical Utility
The receiver operating characteristic (ROC) curves and the areas under the curve (AUCs) were used to measure discriminative ability of the three different models in the training, internal validation and external validation cohorts, respectively. The complex model was also implemented into a nomogram to provide the clinicians an intuitive and quantitative tool for the predictive probability of LNM in NSCLC patients with the increased FDG uptake LNs. Calibration curves were applied to assess the consistency between the nomogram-predicted and actual probability by the Hosmer–Lemeshow goodness-of-fit test in the three cohorts, respectively. A p-value of the Hosmer–Lemeshow test > 0.05 indicated a good fit. In addition, decision curve analysis (DCA) was adopted to determine the clinical usefulness of the nomogram by evaluating the net benefits at different threshold probabilities. The TRIPOD reporting checklist was appended as additional file: Supplementary Material Table 4.

Statistical Analysis
IBM SPSS (version 25.0) and R software (version 3.6.3) were used for data analysis. Numerical data were expressed as mean ± standard deviation, and categorical data were described as numbers. Numerical data: for comparisons between two groups, unpaired t test was performed when the variances were homogeneous, and a non-parametric test (Mann–Whitney) was performed for non-homogeneous variances. For comparisons among three groups, ANOVA was used when the variances were homogeneous. Categorical data: the chi-square test for the analysis of two-by-two contingency tables was performed, and Fisher’s exact test for the analysis of multiple rows and columns was performed. All radiomics features were normalized using the function “scale” of the R-package “base.” Univariable and multivariable logistic regression analyses were performed. The LASSO regression was carried out by the “glmnet” package. The nomogram and calibration curves were performed with the “rms” package. The “pROC” package was used to analyze ROC curves. Comparison of ROCs was performed by the Delong test. Origin (version 2018) was used to plot the ROC curves. DCA was generated using the “rmda” package. All statistical tests were two-sided, and p value < 0.05 was considered statistical significance.

RESULTS

Clinical and Pathological Characteristics
The clinicopathologic characteristics in the training (LN= 159), internal (LN= 69), and external validation cohorts (LN = 60) are summarized in Table 1. There were no statistical differences in gender ($p = 0.572$), age ($p = 0.09$), pathological type ($p = 0.333$), LN station ($p = 0.153$), or pathological acquisition of LNs (surgery/TBNA) ($p= 0.572$) among the three cohorts.

Model 1: PET Radiomics Signature
Due to all extracted features with the inter-reader ICC values >0.8 (mean ± standard deviation, 0.865 ± 0.047), all the 70 original radiomic features were selected for subsequent investigation. Then, the contours drawn by a higher senior physician were finally used to calculate the radiomic features for the modelling.
Subsequently, 70 radiomic candidate features were reduced to 9 with nonzero coefficients by the LASSO regression (Figure 2). Then, by multivariable logistic regression analysis, five features were not statistically significant (p > 0.05). Thus, only four features were used to build the model in the training cohort. The four selected radiomic features were DISCRETIZED_HISTO_Excess Kurtosis, GLRLM_GLNU (Gray-Level Non-Uniformity for run), GLRLM_RLNU (Run Length Non-Uniformity), and NGLDM_Coarseness. Eventually, the four optimal radiomics features were incorporated into radiomics signature. PET_Rad_score = 1.004 + 1.151 × DISCRETIZED_HISTO_Excess Kurtosis − 1.716 × GLRLM_GLNU + 5.651 × GLRLM_RLNU + 1.760 × NGLDM_Coarseness (Table 3). The AUC of the model was 0.820 in the training cohort. The sensitivity, specificity, and accuracy were 61.80%, 90.00%, and 74.22%, respectively (Table 4).

Model 2: CT Image Features and Clinical Data
A comparison of CT image features and clinical data between LNM and non-LNM groups in the training (LNM = 89; non-LNM = 70), internal (LNM = 39; non-LNM = 30), and external validation cohorts (LNM = 33; non-LNM = 27) is listed in Table 2. Univariate analysis demonstrated that there were no statistical differences in age, pathological type, LN station, shape, margin, calcification, and cystic change between the two groups in all three cohorts. Meanwhile, the univariate analysis showed that LN size and CTmean were deemed significantly different between the two groups in all three cohorts. In the training cohort, there also showed a significant statistical difference in CTmax and gender through univariate analyses. Subsequently, gender, LN size, CTmax, and CTmean were finally selected...
through multivariable logistic regression with a stepwise backward selection in the training cohort. The results showed that gender and CTmax were not statistically significant (p > 0.05). Thus, only LN size and CTmean were significantly independent predictive factors. The formula of Model 2 = −1.989 + 0.288 × Size − 0.31 CTmean (Table 3). The AUC of model 2 was 0.780 in the training cohort. The sensitivity, specificity, and accuracy were 79.78%, 68.57%, and 74.84%, respectively (Table 4).

Model 3: Combination of Model 1 and Model 2
A complex model (model 3) incorporating independent predictors from model 1 and model 2 was developed by multivariable logistic regression analysis (Table 3). Calculation formula = 0.6949 + 1.1227 × PET_Rad_score + 0.1114× Size − 0.0531× CTmean. The difference of LN size in Model 2 was statistically significant (p < 0.001; Table 3). However, there was no significant statistical difference in the complex model (p = 0.1659; Table 3). The reason for this phenomenon is that there was a low–moderate correlation between LN size and Rad-score (Pearson correlation coefficient = 0.556, p < 0.001). The discriminatory ability of model 3 displayed the highest with an AUC of 0.874 among the three models in the training cohort (sensitivity 64.05%, specificity 94.29%, and accuracy 77.36%) (Figure 3A and Table 4). Moreover, model 3 was statistically significantly better than both model 1 (p = 0.009) and model 2 (p = 0.011) in the training cohort by the Delong test.

Validation of the Models and Clinical Utility
Based on the formula of the above three models from the training cohort, the AUC values demonstrated that predictive efficacy was calculated in the internal and external validation cohort. The AUC values were 0.785 (model 1), 0.794 (model 2), and 0.845 (model 3) in the internal validation cohort (Figure 3B). The AUC values were 0.808 (model 1), 0.802 (model 2), and 0.841 (model 3) in the external validation cohort (Figure 3C). The sensitivity, specificity, and accuracy are shown in Table 4. Obviously, the complex model showed optimal diagnostic performance among the three cohorts.

### Table 3 | Results of multivariable logistic regression analysis in the three models.

| Models | Included features | Odds ratio (95% CI) | p value | Coefficient | Intercept |
|--------|-------------------|---------------------|---------|-------------|-----------|
| Model 1 | Radiomics signature | DISCRETIZED_HISTO_ExcessKurtosis | 3.160 (1.477–7.228) | 0.004 | 1.151 |
| | | GLRLM_GLNU | 0.180 (0.073–0.395) | <0.001 | -1.716 |
| | | GLRLM_RLNU | 284.479 (26.676–4522.497) | <0.001 | 5.651 |
| | | NGLDM_Coarseness | 5.813 (2.453–15.202) | <0.001 | 1.760 |
| Model 2 | Conventional CT images and clinical data | Size(mm) | 1.333 (1.160–1.533) | <0.001 | 0.298 |
| | | CTmean | 0.969 (0.949–0.990) | 0.003 | -0.31 |
| Model 3: Model 1 + Model 2 | | PET_Rad_score | 3.073 (2.018–4.679) | <0.001 | 1.1227 |
| | | Size (mm) | 1.118 (0.955–1.309) | 0.1659 | 0.1114 |
| | | CTmean | 0.948 (0.926–0.971) | <0.001 | -0.0531 |

PET_Rad_score, PET radomics signature; HISTO, histogram; GLRLM, gray-level run-length matrix; GLNU, gray-level non-uniformity for run; RLNU, run length non-uniformity, NGLDM, neighborhood gray-level different matrix; CI, confidence interval.

The bold values refer to the intercepts for calculation formula.
According to the good diagnostic efficiency of the complex model, a nomogram was generated from the training cohort (Figure 4). The calibration curves demonstrated good consistency between the nomogram-predicted probability of LNM and the actual LNM rate in the three cohorts (Figure 5). There was no statistical significance by the Hosmer–Lemeshow test \( p = 0.437, 0.269, \) and \( 0.531, \) respectively, which showed good fits for the predicted and actual probabilities of LNM in the three cohorts (Figure 5). Finally, the DCA of the nomogram for the training cohort is presented in Figure 6. The decision curve demonstrated that if the threshold probability is more than 10%, using the nomogram to predict LNM adds more benefit than either a “treat all” strategy or a “treat none” strategy.

**DISCUSSION**

With respect to PET/CT, the mediastinal–hilar lymph node which had increased glucose uptake is suspected of having malignant infiltration in patients with NSCLC (7). However, false positivity was frequently detected due to the reactive hyperplasia or granulomatous inflammation (9). In this study, we developed and validated a radiomics nomogram that incorporates PET radiomics signature and two CT image features for the identification of metastatic LNs from the increased FDG uptake LNs in patients with NSCLC.

In recent years, growing evidence suggests that radiomics analysis can better reflect the spatial distribution of voxels and enable profiling of intratumoral heterogeneity in a noninvasive way (15, 25). Radiomics data include first-, second-, and higher-order statistics. Scaling correction was applied to the radiomics values in our study. However, this may not take into account site-by-site differences in scanners, i.e., the radiomics batch effect. Approaches such as the ComBat function are available to address such difficulties with radiomics in follow-up studies. The present study applied radiomics analysis of PET and found that metastatic LNs showed significant differences on DISCRETIZED _HISTO_ExcessKurtosis, GLRLM_GLNU, GLRLM_RLNU, and NGLDM_Coarseness. Among the four texture features, DISCRETIZED _HISTO_ExcessKurtosis is the only feature related to the histogram that represents the distribution of gray level within the VOI. This feature is a relatively simple first-order parameter that describes the heterogeneity of lesions (26). Hu et al. (17) found that histogram features (HISTO_Skewness and HISTO_Kurtosis) extracted from PET had a significant value in differentiating solitary lung adenocarcinoma from tuberculosis. GLRLM_GLNU, GLRLM_RLNU, and NGLDM_Coarseness are the higher-order radiomic features. Several studies have demonstrated that the GLRLM and NGLDM texture analysis can assess the heterogeneity of the lesion (27, 28). Wu et al. (29) reported that the GLRLM textural feature could be useful in differentiating the invasiveness of lung adenocarcinoma. The study by Hoshino et al. (28) presented that NGLDM_Coarseness was correlated with the real miR-1246 expression in the serum of esophageal cancer patients. In the present research, GLRLM_GLNU, GLRLM_RLNU, and NGLDM_Coarseness were

| TABLE 4 | Performance evaluation of three models in three cohorts. |
|--------|------------------|------------------|------------------|
| Model  | Training cohort  | Validation cohort | Validation cohort |
|        | AUC (95% CI)     | Sensitivity (%)  | Specificity (%)  |
| Model 1 | 0.820 (0.754–0.885) | 61.80% | 90.00% | 74.22% |
| Model 2 | 0.780 (0.707–0.853) | 79.78% | 68.57% | 74.94% |
| Model 3 | 0.874 (0.801–0.945) | 84.05% | 77.78% | 77.38% |

AUC, area under the curve; CI, confidence interval.
obviously different between true and false positives of LNM groups, which indicated that metastatic LNs showed high tissue heterogeneity than non-metastatic LNs in patients with NSCLC. Finally, the AUC values of the PET radiomic model were 0.820, 0.785, and 0.808 in the training, internal, and external validation cohorts, respectively, showing good diagnostic efficacy for LNM.

According to previous studies, combining PET glucose metabolic information and CT morphological features is more accurate than PET or CT alone in predicting LN status (30, 31). Thus, the conventional CT imaging features should be considered to further enhance the diagnostic accuracy. Our study demonstrated that there were statistically significant differences in LN size and CTmean between the LNM group and non-LNM group, while there were no statistical differences in shape, margin, calcification, cystic change, and CTmax. However, Zhao et al. showed that shape, margin, calcification, and cystic change had a significant effect on the risk of LNM in patients with thyroid cancer (32). We thought that the discrepancy might be related to different primary lesions, and more importantly the observer’s subjectivity and poor stability of morphological assessment. For the LN size, numerous studies have proved that it is an important parameter in predicting nodal involvement with malignancy, albeit as a relatively unreliable parameter in the assessment of LNs with small lesions (33, 34). Our results also provide evidence that the LN size was a significantly independent predictive factor in the prediction of LNM. For the density of LNs, several studies have shown that the density of benign LNs tends to have higher attenuation, as a result of the chronic granulomatous inflammation (30, 35), which were consistent with our results. For clinical data, our study showed that all of them did not contribute to the model. The possible reason is that a relatively small number of clinical features were included.

The combined model (model 3) was statistically significantly better than both model 1 and model 2 by the Delong test. Obviously, model 3 showed optimal diagnostic performance for predicting the risk of LNM from the increased FDG uptake LNs in NSCLC among the three cohorts, with an AUC of 0.874,
0.845, and 0.841, respectively. LN size in Model 3 had no statistically significant difference, while a statistically significant difference in RAD-score was observed. The results indicated that RAD-score contributed more to the model than size, and the nomogram based on model 3 also proved this. Furthermore, the nomogram demonstrated satisfactory discrimination and calibration in this internal and external validation study, which could be easily available to generate an individual probability of LNM before treatment. Therefore, the nomogram, which can provide a wealth of complementary information of the images, may further extend our knowledge and improve the diagnosis. Although it still could not replace pathological examination at present, the nomogram could help guide sampling of lymph nodes to increase the positive biopsy rates.

Our study still had some limitations. First, this is a retrospective study, thereby having all of the inherent biases of a retrospective study. Second, there was a relatively small sample size and potential selection bias. The main reason is that a definite pathologic result of each hypermetabolic lymph node was necessary, so the lymph node without the result of pathology was excluded. In addition, we have tried our best to match the precise locations of LNs between PET/CT images and surgical resection or EBUS-TBNA and achieve a one-to-one correspondence. Furthermore, most patients who underwent surgery are early stage with metastasis of only a few or even isolated lymph nodes in our study. Around 9% of patients were excluded due to poor image quality, and we do not know how the results will be thrown by the exclusion of qualitatively judged “bad images”. (This will be improved in the future in clinical routine as follows: 1) intravenous catheters were inserted in cubital veins before injection, and a cotton swab is used to avoid bleeding and leakage from the injection site after injection. 2) The patients should be asked to minimize body movements to avoid movement artifacts, and a relatively low dose of sedative drug would be injected if necessary. In addition, the patients should be asked to reduce the flow of respiratory to minimize breathing motion artifacts. 3) We should strengthen the management of image data transfer and storage.) Third, although we have tried our best to match the precise locations of LNs between PET/CT images and surgical resection, it is difficult to completely avoid the matching bias. A further prospective study is therefore needed to solve this problem. Fourth, for VOI, we did not investigate the sensitivity to observer delineation in addition to laborious and time-consuming manual delineation. Fifth, only the texture features of PET images in this study were extracted because hypermetabolic LNs were selected as subjects. Sixth, our study included a relatively small number of clinical parameters in the predictive model. Seventh, selection of hypermetabolic LNs means this diagnostic performance might disappear in a cohort with a moderately active uptake in LNs. Therefore, prospective multicenter studies with large-scale should be carried out to optimize the robustness and reproducibility of the radiomic model.

CONCLUSIONS

In conclusion, our study revealed that the identified PET radiomics signature has the potential to be used as a biomarker for the identification of LNM from the hypermetabolic LNs of
NSCLC. Furthermore, our results provided evidence that a radiomics nomogram that included both the PET radiomics signature and CT images features had a more significant value in identifying true and false positives of mediastinal–hilar LNM detected by PET/CT in patients with NSCLC, which can be used to help clinicians make individual treatment decisions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ethics committee of the First Affiliated Hospital of Wenzhou Medical University and Shaoxing People’s Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin (2018) 68(6):394–420. doi: 10.3322/caac.21492
2. Oser MG, Niederst MJ, Sequist LV, Engelman JA. Transformation From non-Small-Cell Lung Cancer to Small-Cell Lung Cancer: Molecular Drivers and Cells of Origin. Lancet Oncol (2015) 16(4):e165–72. doi: 10.1016/S1470-2045(14)71180-5
3. Tournoy KG, Keller SM, Annema JT. Mediastinal Staging of Lung Cancer: Novel Concepts. Lancet Oncol (2012) 13(5):e221–9. doi: 10.1016/S1470-2045(11)70407-7
4. Rami-Porta R, Call S. Invasive Staging of Mediastinal Lymph Nodes: Mediastinoscopy and Remediastinoscopy. Thorac Surg Clin (2012) 22(2):177–89. doi: 10.1016/j.thorsurg.2011.12.003
5. Gupta NC, Graeber GM, Bishop HA. Comparative Efficacy of Positron Emission Tomography With Fluorodeoxyglucose in Evaluation of Small (<1 cm), Intermediate (1 to 3 cm), and Large (>3 cm) Lymph Node Lesions. Chest (2000) 117(3):773–8. doi: 10.1378/chest.117.3.773
6. Gunluoglu MZ, Melek H, Medetoglu B, Demir A, Kara HV, Dincer SI. The Validity of Preoperative Lymph Node Staging Guidelines of European Society of Thoracic Surgeons in non-Small-Cell Lung Cancer Patients. Eur J Cardiothorac Surg (2011) 40(2):287–90. doi: 10.1016/j.ejcts.2010.11.030
7. Gonzalez-Stawinski GV, Lemaire A, Merchant F, O’Halloran E, Coleman RE, Harpole DH, et al. A Comparative Analysis of Positron Emission Tomography and Mediastinoscopy in Staging Non-Small Cell Lung Cancer. J Thorac Cardiovasc Surg (2003) 126(6):1900–5. doi: 10.1067/mjcs.2003.101036-5
8. Li M, Wu N, Liu Y, Zheng R, Liang Y, Zhang W, et al. Regional Node Staging With 18F-FDG PET-CT in Non-Small Cell Lung Cancer: Additional Diagnostic Value of CT Attenuation and Dual-Time-Point Imaging. Eur J Radiol (2012) 81(8):1886–90. doi: 10.1016/j.ejrad.2011.03.074
9. Antoch G, Stattaus J, Nemat AT, Marnitz S, Beyar T, Kuehl H, et al. Non-Small Cell Lung Cancer: Dual-Modality PET/CT in Preoperative Staging. Radiology (2003) 229(2):526–33. doi: 10.1148/radiol.2292012598
10. Lin WY, Hsu WH, Lin KH, Wang SJ. Role of Preoperative PET-CT in Assessing Mediastinal and Hilar Lymph Node Status in Early Stage Lung Cancer. J Clin Med Assoc (2012) 75(5):203–8. doi: 10.1016/j.jcm.2012.04.004

AUTHOR CONTRIBUTIONS

KT, L-xW, and M-IO contributed to conception and design of the study. Y-rW, Z-hZ, and M-IO collected the clinical information and the imaging data. Q-SD, Y-fZ and Y-rW performed the VOI segmentation. M-IO, KT and LW analyzed the data. M-IO contributed to the drafting and writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Zhejiang Public Welfare Technology Application Research Project, China (LGF21H010009).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.710909/full#supplementary-material
21. Pak K, Kim K, Kim MH, Eom JS, Lee MK, Cho JS, et al. A Decision Tree Model for Predicting Mediastinal Lymph Node Metastasis in non-Small Cell Lung Cancer With F-18 FDG PET/CT. PloS One (2018) 13(2):e0193403. doi: 10.1371/journal.pone.0193403

22. Varela-Lema L, Fernandez-Villar A, Ruano-Ravina A. Effectiveness and Safety of Endobronchial Ultrasound-Transbronchial Needle Aspiration: A Systematic Review. Eur Respir J (2009) 33(5):1156–64. doi: 10.1183/09031936.00097908

23. VanderLaan PA, Wang HH, Majid A, Folch E. Endobronchial Ultrasound-Guided Transbronchial Needle Aspiration (EBUS-TBNA): An Overview and Update for the Cytopathologist. Cancer Cytopathol (2014) 122(8):561–76. doi: 10.1002/cncy.21431

24. Zhao Y, Li X, Li L, Wang X, Lin M, Zhao X, et al. Preliminary Study on the Diagnostic Value of Single-Source Dual-Energy CT in Diagnosing Cervical Lymph Node Metastasis of Thyroid Carcinoma. J Thorac Dis (2017) 9 (11):4758–66. doi: 10.21037/jtd.2017.09.151

25. Decaluwe H, Stanzi A, Dooms C, Fieuws S, Coosemans W, Depypere L, et al. Central Tumour Location Should be Considered When Comparing N1 Upstaging Between Thoracoscopic and Open Surgery for Clinical Stage I Non-Small-Cell Lung Cancer. Eur J Cardiothorac Surg (2016) 50(1):110–7. doi: 10.1093/ejcts/ezy489

26. Wagner F, Hakami YA, Warmock G, Fischer G, Huellner MW, Veit-Haibach P. Comparison of Contrast-Enhanced CT and [(18)F]FDG PET/CT Analysis Using Kurtosis and Skewness in Patients With Primary Colorectal Cancer. Mol Imaging Biol (2017) 19(3):795–803. doi: 10.1007/s11307-017-0166-x

27. Karacavus S, Yilmaz B, Tasdemir A, Kayaalti O, Kaya E, Icer S, et al. Can Laws Be a Potential PET Image Texture Analysis Approach for Evaluation of Tumor Heterogeneity and Histopathological Characteristics in NSCLC? J Digit Imaging (2018) 31(2):210–23. doi: 10.1007/s10278-017-9992-3

28. O’Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. Clin Cancer Res (2015) 21(2):249–57. doi: 10.1158/1078-0432.CCR-14-0990

29. Wu L, Gao C, Xiang P, Zheng S, Pang P, Xu M. CT-Imaging Based Analysis of Invasive Lung Adenocarcinoma Presenting as Ground Glass Nodules Using Peri- and Intra-Nodular Radiomic Features. Front Oncol (2020) 10:838. doi: 10.3389/fonc.2020.00838

30. Kim BT, Lee KS, Shim SS, Choi JY, Kwon OJ, Kim H, et al. Stage T1 non-Small Cell Lung Cancer: Preoperative Mediastinal Nodal Staging With Integrated FDG PET/CT—a Prospective Study. Radiology (2006) 241(2):501–9. doi: 10.1148/radiol.2412051173

31. Darling GE, Maziaik DE, Inculet RI, Gulenchyn KY, Driedger AA, Ung YC, et al. Positron Emission Tomography-Computed Tomography Compared With Invasive Mediastinal Staging in non-Small Cell Lung Cancer: Results of Mediastinal Staging in the Early Lung Positron Emission Tomography Trial. J Thorac Oncol (2011) 6(8):1367–72. doi: 10.1097/JTO.0b013e318226c912

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ouyang, Wang, Deng, Zhu, Zhao, Wang, Wang and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.