Diagnostic Performance of Machine Learning Models Based on \(^{18}\text{F-FDG} \text{PET/CT} \) Radiomic Features in the Classification of Solitary Pulmonary Nodules

Objectives: This study aimed to evaluate the ability of \(^{18}\text{F-FDG} \text{PET/CT} \) radiomic features combined with machine learning methods to distinguish between benign and malignant solitary pulmonary nodules (SPN).

Methods: Data of 48 patients with SPN detected on \(^{18}\text{F-FDG} \text{PET/CT} \) scan were evaluated retrospectively. The texture feature extraction from PET/CT images was performed using an open-source application (LIFEx). Deep learning and classical machine learning algorithms were used to build the models. Final diagnosis was confirmed by pathology and follow-up was accepted as the reference. The performances of the models were assessed by the following metrics: Sensitivity, specificity, accuracy, and area under the receiver operator characteristic curve (AUC).

Results: The predictive models provided reasonable performance for the differential diagnosis of SPNs (AUCs ~0.81). The accuracy and AUC of the radiomic models were similar to the visual interpretation. However, when compared to the conventional evaluation, the sensitivity of the deep learning model (88% vs. 83%) and specificity of the classic learning model were higher (86% vs. 79%).

Conclusion: Machine learning based on \(^{18}\text{F-FDG} \text{PET/CT} \) texture features can contribute to the conventional evaluation to distinguish between benign and malignant lung nodules.

Keywords: Solitary pulmonary nodule, PET/CT, radiomic, machine learning
Introduction

Lung cancer is an important health problem, representing about a quarter of all cancers (1). Early-stage lung cancer may manifest as pulmonary nodules with several distinct features on medical imaging. A solitary pulmonary nodule (SPN) is defined as a well-marginated, rounded parenchymal lesion less than 30 mm in diameter, not associated with other lung pathologies. Common causes of SPN include benign diseases such as infectious granulomas and hamartomas, as well as primary or metastatic lung cancers (2). The management of patients with SPN includes periodic follow-up or further imaging and histopathological examination, considering the malignancy risk (3,4). Positron emission tomography/computed tomography (PET/CT) are widely preferred imaging techniques to detect and characterize SPN, however their diagnostic efficacy does not fully meet clinical needs (5,6).

Radiomics is defined as obtaining high-throughput quantitative features and information from medical images and is a promising approach that has received widespread attention recently (7,8,9,10). Previously, classical machine learning methods and more recently, artificial intelligence applications have been explored for a wide variety of potential uses in lung cancer imaging (11,12,13). Deep learning algorithms using large datasets, such as those from lung cancer screening trials, detect and classify pulmonary nodules with high diagnostic accuracy (13,14). Several predictive models with generally high diagnostic accuracy based on a combination of radiomic features from lung CT and PET/CT have been proposed for different clinical goals (15,16,17,18,19). Preliminary evidence from these studies is promising however more research is needed to verify these results before clinical application. In this study, we aimed to develop predictive models based on 18F-fluorine-fluorodeoxyglucose (18F-FDG) PET/CT texture features for the differential diagnosis of SPN and to evaluate the diagnostic performance of these models.

Materials and Methods

Study Populations

The data of patients who underwent 18F-FDG PET/CT between January 2014 and December 2018 were analyzed retrospectively. The patients included had all the criteria following: (i) 18F-FDG avid SPN detected on PET/CT (n=108); (ii) availability of pathological evidence or at least one-year follow-up (n=80) for the final diagnosis of nodules, as a reference standard. The exclusion criteria are as follows: (i) Nodules at the base of the lungs likely to cause respiratory artifacts (n=15); (ii) nodules with too small metabolic volume to allow adequate tissue features to be extracted (n=17). Finally, the data of 48 patients were evaluated under the above criteria. The Local Ethics Committee of Canakkale Onsekiz Mart University Faculty of Medicine approved this study under the decision number: 09.12.2020/2020-14 and patient informed consent was waived.

PET/CT Acquisition Procedure

18F-FDG PET/CT scans were performed using an integrated PET/CT system (Gemini TF16 PET/CT; Philips Medical Systems). PET images were acquired 60±5 minutes after the intravenous injection of 18F-FDG at a dose of 350-550 MBq in patients who fasted for at least 6 hours and had blood glucose <150 mg/dL. First, a low-dose CT scan (120 kVp peak voltage, of 60-150 mA automated tube current, and 5 mm slice thickness) without contrast enhancement was acquired from the skull vertex to the proximal thigh. Then, PET images were acquired for 2-3 minutes per bed position in 3D mode. PET images were reconstructed using the line-of-response row-action maximum likelihood algorithm (LOR-RAMLA; Philips Astonish TF).

PET/CT Image Interpretation

The PET images were reviewed by two experienced nuclear medicine specialists blinded by the final diagnosis, and the final decision was reached by consensus. The decision for benign and malignant nodules was based on 18F-FDG avidity on PET, along with CT features such as size, margin, density, and calcification (20).
Feature Extraction

An open-source application (LIFEx version 6.30) was used for texture analysis from PET/CT images (21). This application declares Image Biomarker Standardization Initiative compliance. A fixed relative thresholding technique was applied for the tumor delineation on images. A 3-D spherical volume of interest (VOI) was initially placed on the entire lesion. A 40% maximum standardized uptake value (SUV_{max}) threshold was applied to (semi)automatically delineate the VOI of the target lesion on the PET images. All volumes were spatially resampled of 4x4x4 mm in size; absolute resampling was used for intensity rescaling with bounds from 0 to 20 SUV (64 bins, 0.32 fixed bin width); and 64 gray levels were applied for intensity discretization. Radiomic features derived from PET images included conventional indices; first-order features-histogram; shape features; second-order texture features [gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level zone length matrix (GLZLM) and neighborhood gray-tone different matrix (NGLDM)]. A detailed description of the texture parameters can be found at http://www.lifexsoft.org.

Model Establishment

First, feature selection and dimensionality reduction were applied to the feature dataset using the recursive feature elimination (RFE) method. The RFE is a feature selection method that fits a model and removes the weakest features until the specified number of features is reached (22). We build two prediction models based on supervised machine learning classification algorithms selected feature sets: Extreme gradient boosting (XGB) and deep neural network (DNN) to distinguish between benign and malignant nodules. XGB is a tree-based algorithm under the supervised branch of machine learning. XGB, which ensembles the decision tree methods, uses a computationally efficient descent algorithm to minimize errors while adding new trees (19). Deep learning is multi-layer feed-forward neural network that accepts images as input and can be trained end-to-end in a supervised method while learning highly discriminative image features. The opportunity to use large databases has paved the way for the wider adoption of machine/deep learning techniques, particularly in lung cancer assessment (14).

For all models, the dataset was randomly split into two sets using 70% of the samples for training/validating the models and the remaining 30% for testing the results. The models were evaluated using k-fold cross-validation, with three repeats and 10 folds. Figure 1 illustrates the workflow of the radiomic analysis.

Statistical Analysis

We used IBM SPSS statistics software (version 23.0; SPSS Inc.) and Python software to perform statistical analyses. We investigated the performance of predictive models and compared them with the visual evaluation. The following metrics obtained through the confusion matrix were used to compare the performance of the models: Sensitivity, specificity, accuracy, and area under the receiver operator characteristic curve.

Results

In total, the records of 80 patients with SPN were reviewed. Thirty-two patients were excluded under the exclusion criteria. As a result, the study group consisted of 48 patients (31 males, 17 females) with a mean age of 62.38±11.27 years. All of the malignant nodules and 12 of the benign lesions were pathologically proven; the diagnosis of benign lesions was confirmed by follow-up in 5 patients. Thirty-one lesions were malignant nodules, and 17 lesions were benign. The most common malignant diagnosis was adenocarcinoma (58%), while the benign disease was a granulomatous change (53%). The diagnosis and subtypes of SPNs are summarized in Table 1. The majority of malignant nodules (71%) occurred in the upper lobes, whereas about half of the benign nodules (48%) occurred in the lower lobes. Central calcification was observed in four of the benign nodules and punctual calcification was observed in one of the malignant nodules. While most benign nodules tend to have well-defined edges, about half of the malignant nodules have irregular and poorly defined margins. The average diameter of malignant nodules was 20.32 mm (range 16.1-30) and that of benign nodules was 16.9 mm (range 14.2-30). The average SUV_{max} of malignant nodules was 5.46 (range 1.88-10.33) and that of benign nodules was 2.06 (range 1.12-6.77). While SUV_{max} was <2.5 in 24% (4/17) of malignant nodules, SUV_{max} was >2.5 in 23% (7/31) of benign nodules.

The ten most relevant PET features obtained after feature selection and used to develop predictive models are represented in Table 2. The three features with the highest score by the assessment of feature importance were GLZLM_SZLGE (n=30), HISTO_Energy (n=21), and SUV_{bboxmax} (n=21). A few of the second-order features (D_HISTO_Energy, GLCM_Homogeneity, NGLDM_Busyness) were higher in benign nodules, while conventional SUV-related features and other second-order features were higher in the malignant group. Texture features that differ significantly between malignant and benign nodules are shown in Table 3.

Table 4 shows the performance of radiomic models and visual interpretation in the differential diagnosis of SPN.
The overall diagnostic performances of both models were close to each other. The DNN model improved sensitivity, while the XGB model increased specificity compared to visual assessment.

**Discussion**

In this study, we evaluated the performance of machine learning models based on 18F-FDG PET/CT radiomic features for SPN classification. We have shown that the diagnostic accuracy of predictive models is higher than that of commonly used clinical metrics and visual interpretation. The improved diagnostic performance could benefit by

---

**Table 1. The diagnosis and subtypes of the SPNs**

| Type            | Diagnosis                  | Number |
|-----------------|----------------------------|--------|
| Benign (n=17)   | Infectious granuloma       | 9      |
|                 | Hamartoma                  | 3      |
|                 | Not specified*             | 5      |
| Malignant (n=31)| Adenocarcinoma             | 18     |
|                 | Squamous cell carcinoma    | 7      |
|                 | Metastasis                 | 2      |
|                 | Small cell carcinoma       | 2      |
|                 | Carcinoid tumor            | 2      |

*Diagnosed by follow-up, SPNs: Solitary pulmonary nodules

---

**Table 2. The most relevant PET features obtained after feature selection**

| Feature    | Description                  |
|------------|------------------------------|
| f0         | C_SUVmean                    |
| f1         | D_HISTO_Energy               |
| f2         | GLCM_Energy                  |
| f3         | GLCM_Dissimilarity           |
| f4         | GLRLM_HGRE                   |
| f5         | GLRLM_SRHGE                  |
| f6         | GLRLM_LRHGE                  |
| f7         | NGLDM_Busyness               |
| f8         | GLZLM_SZLGE                  |
| f9         | GLZLM_SZHGE                  |

SUV: Standardized uptake value, GLCM: Gray-level co-occurrence matrix, GLRLM: Gray-level run-length matrix, NGLDM: Neighborhood gray-tone different matrix, GLZLM: Gray-level zone length matrix

---

Figure 1. The flowchart of radiomics. 1) The VOI was (semi)automatically defined on PET/CT images with a threshold of 40% of the SUVmax. 2) Radiomic features from the VOI were extracted, including first-order and second-order features. 3) Predictive models were established by feature selection methods and classification methods. 4) The model’s performance was evaluated by appropriate statistical methods. VOI: Volume of interest, PET/CT: Positron emission tomography/computed tomography, SUVmax: Maximum standardized uptake value.
Table 3. Texture features with significant differences between malignant benign nodules

| Features               | Benign SPN | Malign SPN | p       |
|------------------------|------------|------------|---------|
| C_SUV_{mean}           | 1.19±0.75  | 2.61±1.27  | 0.0003  |
| C_SUV_{median}         | 2.52±1.63  | 6.53±2.62  | <0.0001 |
| C_SUV_{maximum}        | 4.11±2.56  | 10.85±4.52 | <0.0001 |
| D_HISTO_Energy         | 0.22±0.14  | 0.08±0.03  | <0.0001 |
| GLCM_Homogeneity       | 0.54±0.17  | 0.33±0.11  | <0.0001 |
| GLCM_Energy            | 0.08±0.19  | 0.01±0.01  | <0.0001 |
| GLCM_Dissimilarity     | 1.27±1.21  | 4.72±2.13  | <0.0001 |
| GLRLM_HGRE             | 124±104    | 567±232    | <0.0001 |
| GLRLM_SRHGE            | 95±121     | 543±423    | <0.0001 |
| GLRLM_LRHGE            | 148.95±137.49 | 672.55±470.7 | <0.0001 |
| NGLDM_Busyness         | 0.87±0.52  | 0.24±0.23  | <0.0001 |
| GLZLM_SZLGE            | 0.01±0.013 | 0.004±0.008 | <0.0001 |
| GLZLM_SZHGE            | 48.75±45.77 | 342.91±306.06 | <0.0001 |

SPN: Solitary pulmonary nodule, SUV: Standardized uptake value, GLCM: Gray-level co-occurrence matrix, GLRLM: Gray-level run-length matrix, NGLDM: Neighborhood gray-tone different matrix, GLZLM: Gray-level zone length matrix

Table 4. The estimated performances of radiomic models and visual interpretation

| Model  | SN   | SP   | ACC  | AUC  |
|--------|------|------|------|------|
| DNN    | 88%  | 82%  | 0.80 | 0.81 |
| XGB    | 81%  | 86%  | 0.79 | 0.80 |
| VI     | 83%  | 79%  | 0.80 | 0.80 |

SN: Sensitivity, SP: Specificity, ACC: Accuracy, AUC: Area under the curve, DNN: Deep neural network, XGB: Extreme gradient boosting, VI: Visual interpretation

PET/CT radiomics in lung cancer have been investigated for clinical goals such as characterization of nodules, histological subtyping, prediction of survival, and response to therapy (11,12). Few studies that focused on the characterization of pulmonary nodules demonstrated the ability of PET/CT radiography to distinguish between malignant and benign lesions (15,16,17,18,19,25,26,27). In the studies, the results of machine learning models trained with texture features derived from 18F-FDG PET/CT were compared with standard metrics [SUV, metabolic tumor volume (MTV), and total lesion glycolysis] and/or visual interpretation evaluation. Studies with dual time point 18F-FDG PET/CT, particularly the results obtained with tissue properties in delayed images, provided important improvements for classifying SPNs (14,15,27,28). Texture features that reflect intra-lesional heterogeneity, termed second-order texture features in this study, showed significant differences between the malignant and benign groups, as reported in studies.

Our predictive models showed reasonable diagnostic performance with balanced sensitivity and specificity for the differential diagnosis of SPNs. Compared with the conventional evaluation results, the deep learning model increased sensitivity, while the classic machine learning model increased specificity. The overall performance of our models was consistent with the results of the cited studies; however, the improvement in diagnostic accuracy was less than the reported results (15,16,17,18,19). This difference may be due to the small size of our cohort and the fact that the diagnosis of all nodules was not confirmed by pathology. Additionally, most investigators created models with tissue features from dual time-point PET/CT, and higher diagnostic accuracy was reported, particularly from delayed images.

In standard PET/CT scans, respiratory motion adversely affects both alignment and image sharpness, resulting in reduced tracer uptake and an overestimation of MTV (29). Several PET/CT radiomics articles have reported that respiratory motion significantly affects the values of texture features of lung lesions (30,31). These effects differ according to the location of the lesion in the lung; for example, it is more prominent in the lower lobes. Therefore, nodules located in the lower lobes of the lungs were excluded from the radiomic analysis in our study.

It is difficult to compare the results of machine learning studies reported on PET/CT imaging of lung cancer, as researchers have chosen different materials and methods to construct their models. We performed PET/CT radiomic analysis with two models based on classifiers and feature selection methods to improve the quality score of our study, as suggested by Lambin et al. (32). Zhou et al. (19) compared the performance of machine learning models based on PET/CT radiomics for the classification of lung lesions (16). They reported that most classifiers combined with appropriate feature selection methods showed excellent discrimination. They suggested that gradient
boosting decision tree and random forest are the best classification methods. In another study, the deep learning method was compared with classical machine learning methods to classify mediastinal lymph node metastasis in PET/CT images (33). The authors reported that there was no significant difference between the results of deep learning and classical methods, however, machine learning methods have higher sensitivity but lower specificity than doctors.

**Study Limitations**

Several limitations should be considered in our study. First, this study was a retrospective analysis and inherent selection bias existed. Secondly, the small size of our study population may have adversely affected the performance of machine learning algorithms. Thirdly, the study’s lack of external validation limits the generalizability of our results.

**Conclusion**

In this study, we performed a machine learning-based analysis of pulmonary nodules using PET/CT images. We found that 18F-PET/CT-based radiomic features can provide added value in differentiating SPNs. The method should be further confirmed in large-scale multicenter, ideally prospective studies so that it can be applied in routine clinical practice.

**Ethics**

**Ethics Committee Approval:** The Local Ethics Committee of Canakkale Onsekiz Mart University Faculty of Medicine approved this study under the decision number: 09.12.2020/2020-14.

**Informed Consent:** Patient informed consent was waived.

**Peer-review:** Externally peer-reviewed.

**Authorship Contributions**

Design: Y.S.S., Data Collection or Processing: R.U.E., B.A.P., Y.S.S., Ç.U., T.T.E., H.O.T., Analysis or Interpretation: Y.S.S., S.O., R.U.E., Literature Search: Y.S.S., R.U.E., Writing: Y.S.S., S.O.

**Conflict of interest:** No conflict of interest was declared by the authors.

**Financial Disclosure:** The authors declare that this study has received no financial support.

**References**

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.
2. Nasim F Ost DE. Management of the solitary pulmonary nodule. Curr Opin Pulm Med 2019;25:344-353.
21. Nioche C, Orthac E, Boughdad S, Reuzé 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.

22. Jeon H, Oh S. Hybrid-recursive feature elimination for efficient feature selection. Applied Sciences 2020;10:3211.

23. Taralli S, Scolozzi V, Fofi M, Ricciardi S, Forcione AR, Cardillo G, Calcagni ML. 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.

24. Deppen SA, Blume JD, Kensinger CD, Morgan AM, Aldrich MC, Massion PP, Walker RC, McPheeters ML, Putnam JB Jr, Grogan EL. Accuracy of FDG-PET to diagnose lung cancer in areas with infectious lung disease: a meta-analysis. JAMA 2014;312:1227-1236.

25. Du D, Gu J, Chen X, Lu W, Feng Q, Rahmim A, Wu H, Lu L. Integration of PET/CT radiomics and semantic features for differentiation between active pulmonary tuberculosis and lung cancer. Mol Imaging Biol 2021;23:287-298.

26. Hu Y, Zhao X, Zhang J, Han J, Dai M. Value of 18F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis. Eur J Nucl Med Mol Imaging 2021;48:231-240.

27. Nakajo M, Jingui M, Aoki M, Tani A, Sato M, Yoshiura T. The clinical value of texture analysis of dual-time-point 18F-FDG-PET/CT imaging to differentiate between 18F-FDG-avid benign and malignant pulmonary lesions. Eur Radiol 2020;30:1759-1769.

28. Teramoto A, Tsujimoto M, Inoue T, Tsukamoto T, Imaizumi K, Toyama H, Saito K, Fujita H. Automated classification of pulmonary nodules through a retrospective analysis of conventional CT and two-phase PET images in patients undergoing biopsy. Asia Ocean J Nucl Med Biol 2019;7:29-37.

29. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. Radiother Oncol 2012;102:239-245.

30. Oliver JA, Budzевич M, Zhang GG, Dilling TJ, Latifi K, Moros EG. Variability of image features computed from conventional and respiratory-gated PET/CT images of lung cancer. Transl Oncol 2015;8:524-534.

31. Grootjans W, Tixier F, van der Vos CS, Viens D, Le Rest CC, Bussink J, Oyen WJ, de Geus-Oei LF, Visvikis D, Visser EP. The impact of optimal respiratory gating and image noise on evaluation of intratumor heterogeneity on 18F-FDG PET imaging of lung cancer. J Nucl Med 2016;57:1692-1698.

32. Lambin P, Leijenaar RTH, Desl TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AIG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-762.

33. Wang H, Zhou Z, Li Y, Chen Z, Lu P, Wang W, Liu W, Yu L. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. EJNMMI Res 2017;7:11.