DOI: 10.55415/deep-2022-0043.v1

# Advances in radiomics in the differential diagnosis of pulmonary nodules

Li Yi，Minjing Zuo*

Department of Radiology, The Second Affiliated Hospital of Nanchang University, Jiangxi, China.
* Correspondence: Minjing Zuo
Phone Number: 15970667423. E-mail: 124542022@qq.com. Address: No.1 Minde Road, Donghu District, Nanchang City, Jiangxi Province. Postcode: 330006.
This is a funded project of Jiangxi Provincial Education Department (No. GJJ200106) and Jiangxi Provincial Science and Technology Department Applied Research Incubation Program (No. 20212BAG70048)

According to cancer statistics[1, 2], lung cancer is the most common and deadly cancer worldwide, mainly because the vast majority of patients have no obvious clinical symptoms and are already in an advanced stage of the disease at initial diagnosis. Low-dose computed tomography (LDCT) screening can help detect early-stage lung lesions and has reduced lung cancer mortality by 20%[3]. However, this has led to an increasing number of lung nodules being detected, including a large number of false-positive cases, with more than 95% of the nodules found being benign [4], and it is extremely difficult to make a differential diagnosis and perform various treatment measures for these nodules. The Fleischner Society guidelines and the Lung CT Screening Reporting and Data System (Lung- RADS) recommend that the size and growth pattern be closely observed when pulmonary nodules are detected[5, 6]. However, awareness of the recommendations and treatment decisions in clinical

DOI: 10.55415/deep-2022-0043.v1

practice vary between radiologists and pulmonologists [7], and intensive long-term follow-up CT may increase radiation dose to patients and cause psychological problems for patients. The spread of lung cancer screening and the development of treatment modalities urgently require good noninvasive biomarkers that can accurately diagnose, classify, and risk stratify screening and incidentally detected lung nodules. This is made possible by the development of imaging histology as a new technology that identifies, extracts, quantifies, and analyzes imaging features from imaging images to better characterize the phenotype of pulmonary nodules than is possible with the naked eye. The focus of this article is an overview of research advances in imaging histology in the differential diagnosis of pulmonary nodules.

## 1. Concept and basic process of radiomics

The concept of radiomics was first proposed by Dutch scholar Lambin in 2012 [8] and aims at high-throughput extraction of objective and quantitative hard-to-identify radiological features from image data, which are then transformed into high-dimensional data information. In this way, image-to-data conversion is achieved, bridging the gap between image data and precision medicine by enabling broader and deeper mining, prediction, and analysis of bulk data to help physicians make more accurate diagnoses [9].

### 1.1 Acquisition of image data and Segmentation of region of interest

DOI: 10.55415/deep-2022-0043.v1

Collecting image data of interest as a study object is most commonly done with CT images, and in recent years imaging histology is increasingly used for other imaging modalities (e.g., MRI, PET, ultrasound, etc.). The main challenge is that there are no uniform standards for image acquisition and processing in different institutions. Different images vary widely in terms of, for example, scanning equipment, radiation dose, acquisition protocol, pixel size, slice thickness, use of contrast agents, and post-processing parameters[10]. Therefore, standardization principles must be followed in image acquisition to minimize the disruptive effects of image differences [11]. Image segmentation refers to the delineation of the lesion or area of interest to be examined and includes three methods of implementation: manual, semi-automatic, or automatic. Manual segmentation is time-consuming and has uncertain reproducibility, while automatic segmentation is less accurate in some specific areas of nodules (hilum, pleura, mediastinum, etc.). The semi-automatic method, on the other hand, compensates better for the first two points and is more widely used [12].

**1.2 Feature extraction and selection** Radiomics feature extraction is often based on various commercial or open source software packages. The main features include histogram features, shape features, intensity features, texture features, and features extracted by various filters. Histogram features include grayscale mean, maximum, minimum,

DOI: 10.55415/deep-2022-0043.v1

variance, percentile, etc. Texture features include grayscale co-occurrence matrix, grayscale distance matrix, grayscale region matrix, etc. Shape features include diameter, perimeter, area, volume, etc. The number of these features is often huge and there are many irrelevant and redundant features. To avoid dimensionality disasters, feature selection and dimensionality reduction methods (such as linear regression, recursive feature elimination, cluster analysis, and principal component analysis) can select the most important, repeatable, non-redundant, and relevant features for later model construction[13].

**1.3 Model construction and evaluation** An accurate, reliable, and efficient model is a key factor in the success of radiomics. Common model analysis methods can be divided into unsupervised and supervised data analysis. The former include k-means clustering, hierarchical clustering, and consistent clustering, while the latter include logistic regression, Cox regression, random forests, support vector machines, and artificial neural networks. The latter is more commonly used to train models for predicting unknown data based on data with outcomes, but current research does not define the optimal method but depends on the actual classification application [14].

## 2. Advantages of radiomics for differential diagnosis of pulmonary nodules

Traditional methods of assessing pulmonary nodules include CT, MRI,

DOI: 10.55415/deep-2022-0043.v1

and F-18 fluorodeoxyglucose (FDG) PET. High-resolution CT and target scans can provide information about nodule shape, borders, and internal vascularand bronchial structures. The advantages of this noninvasive, efficient, and inexpensive method make it the most widely used imaging examination tool. However, it is difficult to characterize different imaging modalities by visual assessment, and it is susceptible to human factors with only moderate agreement[15]. MRI has the merits of multiparametric imaging, but its lack of resolution and respiratory artifacts greatly affect it [16]. PET The scan reflects the metabolic information of the lesion and has high sensitivity for the diagnosis of benign and malignant lung nodules, but it is limited by its resolution and the inert state of some smaller lung cancer nodules, with no discriminatory effect in nodules smaller than 10 mm [17].

Radiomics can tap more image information and integrate the advantages of multiple diagnostic methods to effectively improve the correct diagnosis of pulmonary nodules. Liu et al[18] retrospectively collected and analyzed CT images of 875 patients with pulmonary nodules and established a radiomics model that can predict the benignity and malignancy of pulmonary nodules well. The study by Chen et al[19] also demonstrated the excellent performance of the radiomics model. Garau et al [20] validated the radiomics model using an external validation cohort to evaluate the generalization performance and showed that both

DOI: 10.55415/deep-2022-0043.v1

support vector machine and neural network models could accurately identify benign and malignant lung nodules with AUC values >of 0.89 in both the coordinated and uncoordinated cohorts of the training set. The above study demonstrated that the radiomics-based prediction model has very good accuracy and great potential as a noninvasive tool for preoperative prediction of benign and malignant lung nodules.

Traditional models for assessing pulmonary nodules often include CT semantic signs, clinical laboratory factors, and so on. To investigate how the predictive performance of radiomics compares to traditional models, Jing et al[21] used unenhanced CT images of patients with pulmonary nodules for radiomics analysis and combined imaging histologic features and clinical factors to create a combined model for the diagnosis of benign and malignant pulmonary nodules with an AUC of 0.940 and a 95% confidence interval of 0.883-0.998, outperforming both the radiomics model alone and the model with clinical factors and demonstrating significantly higher accuracy than the Mayo Clinic model. Choi et al[22] used an imaging histology model constructed with a support vector machine (SVM) classifier to predict benign and malignant pulmonary nodules preoperatively and achieved an accuracy of 84.6%, which was 12.4% higher than that of the American College of Radiology Lung CT Screening Report and Data System (Lung- RADS). In the field of pulmonary nodule diagnosis, radiomics can combine several diagnostic

DOI: 10.55415/deep-2022-0043.v1

advantages of imaging and clinical practice to achieve higher predictive performance compared with conventional assessment models.

Mao et al[23] evaluated the performance of radiomics in the differential diagnosis of pulmonary nodules of different sizes to identify benign and malignant isolated small pulmonary nodules with a diameter of 6-15 mm, and the prediction accuracy of the model reached 89.8%, which was better than that of Lung- RADS. For the identification of benign and malignant pulmonary nodules less than 10 mm in diameter, radiological models based on logistic regression[24] and random forest[25] achieved excellent performance, while Xu et al[26] divided 373 patients with collected pulmonary nodules into three groups according to size: T1a (0-1 cm), T1b (1 - 2 cm), and T1c (2 - 3 cm). The radiological characteristics were studied separately and models were constructed to predict the benignity and malignancy of the nodules. The results showed that the T1a group had the best predictive performance (AUC of 0.84: 0.78: 0.79). This indicates that radiomics takes advantage of the high resolution of CT and the evaluation performance is not affected by the size of the nodule, which is still applicable in predicting benign and malignant nodules in small lung nodules.

**3. Application of radiomics in the differential diagnosis of benign and malignant pulmonary nodules**

Most solid lung nodules detected on chest examination CT are benign,

DOI: 10.55415/deep-2022-0043.v1

about 80% are granulomas[27], and it is difficult to identify them with lung cancer, especially adenocarcinoma of the lung. Both have great similarities in the site of origin, size, density, CT signs, and even PET presentation. In contrast, the radiological model created by Orooji et al[28] based on non-enhanced images from CT was able to distinguish lung adenocarcinoma well from sarcoidosis (AUC value of 0.77) and outperformed manual review by an experienced thoracic radiologist (0.69) and thoracic surgeon (0.72). In another study, [29]an external validation cohort from different medical institutions was used for evaluation, and the bar graph prediction model constructed based on 22 radiological characteristics, gender, and fractal shape had AUCs of 0.885 and 0.808 in the training and external validation datasets, respectively, which facilitated the individualized preoperative diagnostic treatment of patients with pulmonary nodules and confirmed the generalizability of the model. As part of granulomatous lesions, cryptococcal infections caused by Cryptococcus novelis or Cryptococcus gattii, which appear as an isolated nodular pattern, are difficult to distinguish from lung cancer by CT semantic features, clinical presentation, and laboratory tests[30] Li et al[31] retrospectively collected 296 asymptomatic patients with negative blood tests for tumor markers and fungal markers and could not definitively CT patients with cryptococcosis and lung cancer, who were diagnosed scanned, constructed local deep-learning models (which contained only

DOI: 10.55415/deep-2022-0043.v1

nodal information), local-global deep-learning models (which contained nodal and lung conditions), and radiomicroscopy models, with local-global deep-learning models having the best performance (AUC= 0.88) in distinguishing nodal cryptococcosis from lung cancer. Atypical tuberculomas often have no particular signs such as calcification and satellite foci, but show malignant CT signs such as ridge, lobulation, pleural depression, etc., which are difficult to distinguish from lung cancer[32]. Feng et al[33] segmented nodules and extracted features from conventional CT images of 426 patients with tuberculous sarcoidosis and lung cancer, which appeared as isolated solid nodules. Three independent predictors of radiologic features, age, and burr sign were used to create a predictive model for a columellar map that demonstrated better diagnostic accuracy (AUC=0.966) than any single model for accurately distinguishing lung cancer from tuberculoma preoperatively. In contrast, another study[34] used convolutional neural networks to extract deep learning features (DLS) from patients with lung cancer and tuberculous sarcoidosis, and found that DLS, gender, age, and fractal shape were independent predictors and were used to generate deep learning columnar line graphs with areas under the curve of 0.889, 0.879, and 0.809 for the training, internal validation, and external validation cohorts, respectively, which outperformed the radiomics model and the clinical radiology model.

DOI: 10.55415/deep-2022-0043.v1

The peritumor region represents a dynamic and complex environment of cellular and noncellular components surrounding the tumor and has been shown to play an important role in tumor biological behavior such as cell migration, inflammation, angiogenesis, and aggressiveness[35]. As part of the lesion, it represents heterogeneity of the lesion that is difficult to detect with conventional imaging techniques and the human eye. Radiomics translates this heterogeneity into quantifiable data types for analysis. Beig et al[36] characterized the heterogeneity of lung cancer and sarcoidosis using radiomic features of different areas of the peritumoral regions. Adenocarcinomas were found to express more low-frequency Gabor features corresponding to their pathologically dense peritumoral tumor-infiltrating lymphocytes, with a smoother texture at CT. In contrast, low-frequency Gabor expression in the peri-sarcoma region corresponds to normal lung tissue with few giant cell nodules in the peritumor region. The study by Calheiros et al[37] also highlighted the role of the perinodular region in distinguishing lung cancer from sarcoidosis, and the radiological model incorporating the features of the perinodular region was improved in all assessment metrics (AUC of 0.916, accuracy of 84.26%, sensitivity of 84.45%, and specificity of 83.84%). Alilou et al[38] constructed a radiological scoring system for nodal margins and interface sharpness (NIS). The AUC of the model improved from 0.77 to 0.84 after intra-nodal shape and texture features were included, and this study

DOI: 10.55415/deep-2022-0043.v1

showed that NIS correctly reclassified 46% of lung nodules diagnosed as suspiciously malignant by Lung- RADS but pathologically benign.

In addition to CT, which is most commonly used as the primary source for radiomics, many studies in recent years have focused on the application of radiomics to other imaging materials. Multi-parametric imaging of MRI provides more information about lesions and has good resolution in soft tissue. A[39] radiomics analysis based on MRI T2WI images showed that models based on 3D features were superior to 2D features in predicting benign and malignant solid lung nodules, with AUC values of more than 0.7 in several combined machine learning methods, and the best models (AUC=0.84) were generated by recursive feature elimination and support vector bodies. Wang et al[40] obtained the best results using T1-weighted, T2-weighted, and apparent diffusion coefficient (ADC) multiparametric MRI images to extract radiological features, and the combined model with multiple MRI parameters showed the highest performance in the test group (AUC 0.88; sensitivity 83%; accuracy 82%; specificity 79%). Fluorodeoxyglucose positron emission tomography (FDG-PET) is extremely sensitive and reflects information about lesion metabolism. Du et al[41] performed a radiological analysis of PET-CT in 172 patients with pathologically confirmed pulmonary tuberculomas and lung cancer and introduced conventional CT semantic features. The predictive model consisting of the combination of both

DOI: 10.55415/deep-2022-0043.v1

features was superior to the radiomic model and the model with the semantic features CT alone (AUC 0.97: 0.94: 0.91). Chen et al[42] used the neighborhood gray difference matrix (NGTDM) as a texture feature on dual time point images (DTPI) PET-CT to discriminate the benignity and malignancy of isolated solid nodules, and the results showed that the radiomic model with delayed period PET-CT was a good predictor of benignity and malignancy of isolated solid nodules, which was superior to the conventional radiomic and clinical models and SUV indicators.

In summary, in the differential diagnosis of pulmonary nodules, from general benign and malignant differentiation to more detailed classification and differentiation of lung cancer nodules from granulomatous lesions, tuberculomas, and cryptococcal infections, from purely imaging histologic features to more comprehensive models that include traditional CT signs, clinical manifestations, and laboratory tests, from tumor region to peritumor region analysis, and from CT to more imaging data such as MRI and PET, imaging histology has shown great potential to effectively aid in the evaluation and treatment of lung nodules.

## 4. Challenges and Prospects

CT radiomics has achieved many results in the differential diagnosis of pulmonary nodules, but as a new technology, there are still some challenges in some aspects: (1) standardization of imaging data, in

DOI: 10.55415/deep-2022-0043.v1

practice in clinical work, there is some variability between devices and parameters between different institutions and devices, which may have an impact on the results. (2) Reproducibility of lesion segmentation. In the present study, most of the segmentation was performed manually, introducing a human factor into the computerized process of radiomicroscopy, and the reproducibility of this process is questionable. (3) Lack of prospective, multicenter studies with large samples, so numerous models have been created but have not been widely accepted.

Despite the above challenges, radiomics has seen an increase in research zeal as an emerging technology that intersects and integrates multiple disciplines and fields. It is anticipated that with the process of globalization, the deepening of communication and collaboration among regions, and the development of computer networks, the acquisition of multicenter, standardized data with large samples and the application of reproducible automatic segmentation algorithms will become possible, and radiomics will improve efficiency, increase diagnostic accuracy, and support precision medicine on a larger scale.

References

DOI: 10.55415/deep-2022-0043.v1

[1] Siegel R L, Miller K D, Jemal A. Cancer statistics, 2019[J]. CA Cancer J Clin, 2019,69(1):7-34.

[2] Chen W, Zheng R, Baade P D, et al. Cancer statistics in China, 2015[J]. CA Cancer J Clin, 2016,66(2):115-132.

[3] de Koning H J, van der Aalst C M, de Jong P A, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial[J]. N Engl J Med, 2020,382(6):503-513.

[4] McWilliams A, Tammemagi M C, Mayo J R, et al. Probability of cancer in pulmonary nodules detected on first screening CT[J]. N Engl J Med, 2013,369(10):910-919.

[5] MacMahon H, Naidich D P, Goo J M, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017[J]. Radiology, 2017,284(1):228-243.

[6] Callister M E, Baldwin D R, Akram A R, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules[J]. Thorax, 2015,70 Suppl 2:i1-i54.

[7] Mets O M, de Jong P A, Chung K, et al. Fleischner recommendations for the management of subsolid pulmonary nodules: high awareness but limited conformance - a survey study[J]. Eur Radiol, 2016,26(11):3840-3849.

[8] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis[J]. Eur J Cancer, 2012,48(4):441-446.

[9] Lambin P, Leijenaar R, Deist T M, et al. Radiomics: the bridge between medical imaging and personalized medicine[J]. Nat Rev Clin Oncol, 2017,14(12):749-762.

[10] Lee G, Bak S H, Lee H Y. CT Radiomics in Thoracic Oncology: Technique and Clinical Applications[J]. Nucl Med Mol Imaging, 2018,52(2):91-98.

[11] Avanzo M, Stancanello J, El N I. Beyond imaging: The promise of radiomics[J]. Phys Med, 2017,38:122-139.

[12] Gillies R J, Kinahan P E, Hricak H. Radiomics: Images Are More than Pictures, They Are Data[J]. Radiology, 2016,278(2):563-577.

[13] Griffié J, Shannon M, Bromley C L, et al. A Bayesian cluster analysis method for single-molecule localization microscopy data[J]. Nat Protoc, 2016,11(12):2499-2514.

[14] Parekh V, Jacobs M A. Radiomics: a new application from established techniques[J]. Expert Rev Precis Med Drug Dev, 2016,1(2):207-226.

[15] van Riel S J, Sánchez C I, Bankier A A, et al. Observer Variability for Classification of Pulmonary Nodules on Low-Dose CT Images and Its Effect on Nodule Management[J]. Radiology, 2015,277(3):863-871.

[16] Ohno Y, Kauczor H U, Hatabu H, et al. MRI for solitary pulmonary nodule and mass assessment: Current state of the art[J]. J Magn Reson Imaging, 2018,47(6):1437-1458.

[17] Nomori H, Watanabe K, Ohtsuka T, et al. Evaluation of F-18 fluorodeoxyglucose (FDG) PET scanning for pulmonary nodules less than 3 cm in diameter, with special reference to the CT images[J]. Lung Cancer, 2004,45(1):19-27.

[18] Liu A, Wang Z, Yang Y, et al. Preoperative diagnosis of malignant pulmonary nodules in lung cancer screening with a radiomics nomogram[J]. Cancer Commun (Lond), 2020,40(1):16-24.

[19] Chen C H, Chang C K, Tu C Y, et al. Radiomic features analysis in computed tomography images of lung nodule classification[J]. PLoS One, 2018,13(2):e192002.

[20] Garau N, Paganelli C, Summers P, et al. External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis[J]. Med Phys, 2020,47(9):4125-4136.

DOI: 10.55415/deep-2022-0043.v1

[21] Jing R, Wang J, Li J, et al. A wavelet features derived radiomics nomogram for prediction of malignant and benign early-stage lung nodules[J]. Sci Rep, 2021,11(1):22330.

[22] Choi W, Oh J H, Riyahi S, et al. Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer[J]. Med Phys, 2018,45(4):1537-1549.

[23] Mao L, Chen H, Liang M, et al. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening[J]. Quant Imaging Med Surg, 2019,9(2):263-272.

[24] Liu Q, Huang Y, Chen H, et al. The development and validation of a radiomic nomogram for the preoperative prediction of lung adenocarcinoma[J]. BMC Cancer, 2020,20(1):533.

[25] Liu Q, Huang Y, Chen H, et al. Computed Tomography-Based Radiomic Features for Diagnosis of Indeterminate Small Pulmonary Nodules[J]. J Comput Assist Tomogr, 2020,44(1):90-94.

[26] Xu Y, Lu L, E L N, et al. Application of Radiomics in Predicting the Malignancy of Pulmonary Nodules in Different Sizes[J]. AJR Am J Roentgenol, 2019,213(6):1213-1220.

[27] Sánchez M, Benegas M, Vollmer I. Management of incidental lung nodules <8 mm in diameter[J]. J Thorac Dis, 2018,10(Suppl 22):S2611-S2627.

[28] Orooji M, Alilou M, Rakshit S, et al. Combination of computer extracted shape and texture features enables discrimination of granulomas from adenocarcinoma on chest computed tomography[J]. J Med Imaging (Bellingham), 2018,5(2):24501.

[29] Chen X, Feng B, Chen Y, et al. A CT-based radiomics nomogram for prediction of lung adenocarcinomas and granulomatous lesions in patient with solitary sub-centimeter solid nodules[J]. Cancer Imaging, 2020,20(1):45.

[30] Chang C C, Sorrell T C, Chen S C. Pulmonary Cryptococcosis[J]. Semin Respir Crit Care Med, 2015,36(5):681-691.

[31] Li S, Zhang G, Yin Y, et al. One deep learning local-global model based on CT imaging to differentiate between nodular cryptococcosis and lung cancer which are hard to be diagnosed[J]. Comput Med Imaging Graph, 2021,94:102009.

[32] Pusch T, Pasipanodya J G, Hall R N, et al. Therapy duration and long-term outcomes in extra-pulmonary tuberculosis[J]. BMC Infect Dis, 2014,14:115.

[33] Feng B, Chen X, Chen Y, et al. Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule[J]. Eur J Radiol, 2020,128:109022.

[34] Feng B, Chen X, Chen Y, et al. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas[J]. Eur Radiol, 2020,30(12):6497-6507.

[35] Mittal V, El R T, Narula N, et al. The Microenvironment of Lung Cancer and Therapeutic Implications[J]. Adv Exp Med Biol, 2016,890:75-110.

[36] Beig N, Khorrami M, Alilou M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas[J]. Radiology, 2019,290(3):783-792.

[37] Calheiros J, de Amorim L, de Lima L L, et al. The Effects of Perinodular Features on Solid Lung Nodule Classification[J]. J Digit Imaging, 2021,34(4):798-810.

[38] Alilou M, Prasanna P, Bera K, et al. A Novel Nodule Edge Sharpness Radiomic Biomarker Improves Performance of Lung-RADS for Distinguishing Adenocarcinomas from Granulomas on Non-Contrast CT Scans[J]. Cancers (Basel), 2021,13(11).

DOI: 10.55415/deep-2022-0043.v1

[39] Wan Q, Zhou J, Xia X, et al. Diagnostic Performance of 2D and 3D T2WI-Based Radiomics Features With Machine Learning Algorithms to Distinguish Solid Solitary Pulmonary Lesion[J]. Front Oncol, 2021,11:683587.

[40] Wang X, Wan Q, Chen H, et al. Classification of pulmonary lesion based on multiparametric MRI: utility of radiomics and comparison of machine learning methods[J]. Eur Radiol, 2020,30(8):4595-4605.

[41] Du D, Gu J, Chen X, et al. Integration of PET/CT Radiomics and Semantic Features for Differentiation between Active Pulmonary Tuberculosis and Lung Cancer[J]. Mol Imaging Biol, 2021,23(2):287-298.

[42] Chen S, Harmon S, Perk T, et al. Using neighborhood gray tone difference matrix texture features on dual time point PET/CT images to differentiate malignant from benign FDG-avid solitary pulmonary nodules[J]. Cancer Imaging, 2019,19(1):56.