A CT-based radiomics integrated model for discriminating pulmonary cryptococcosis granuloma from lung adenocarcinoma—a diagnostic test

Bin Hu[1,2#], Wei Xia[3#], Sirong Piao[1#], Ji Xiong[2,4], Ying Tang[1], Hong Yu[5], Guangyu Tao[5], Linlin Sun[5], Minhui Shen[6], Ajay Wagh[7], Timothy J. Jaykel[8], Ding Zhang[9*], Yuxin Li[1,2*], Li Zhu[5*^]

[1]Department of Radiology, Huashan Hospital, Fudan University, Shanghai, China; [2]Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai, China; [3]Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China; [4]Department of Pathology, Huashan Hospital, Fudan University, Shanghai, China; [5]Department of Radiology, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China; [6]College of Medical Instrument, Shanghai University of Medicine & Health Sciences, Shanghai, China; [7]Section of Pulmonary and Critical Care Medicine/Interventional Pulmonology, The University of Chicago, Chicago, IL, USA; [8]Department of Radiology, Mayo Clinic, Rochester, MN, USA; [9]Department of Pulmonary and Critical Care Medicine, Huashan Hospital, Fudan University, Shanghai, China

Contributions: (I) Conception and design: Y Li, L Zhu; (II) Administrative support: Y Li, L Zhu; (III) Provision of study materials or patients: H Yu, G Tao, L Sun, D Zhang, Y Li, L Zhu, M Shen; (IV) Collection and assembly of data: B Hu, S Piao, Y Tang, J Xiong; (V) Data analysis and interpretation: W Xia, J Xiong; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work and should be considered as co-first authors.

[*]These authors contributed equally to this work for the senior authorship.

Correspondence to: Li Zhu, MD, PhD. Department of Radiology, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, 241 West Huaihai Road, Shanghai 200030, China. Email: augjuly@aliyun.com; Yuxin Li, MD, PhD. Department of Radiology, Huashan Hospital, Fudan University, No. 12 Middle Wulumuqi Road, Jingan District, Shanghai 200040, China; Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai, China. Email: liyuxin@fudan.edu.cn; Ding Zhang, MD, PhD. Department of Pulmonary and Critical Care Medicine, Huashan Hospital, Fudan University, No. 12 Middle Wulumuqi Road, Jingan District, Shanghai 200040, China. Email: zhangding12@fudan.edu.cn.

Background: Chest computed tomography (CT) is a critical tool in the diagnosis of pulmonary cryptococcosis as approximately 30% of normal immunity individuals may not exhibit any significant symptoms or laboratory findings. Pulmonary cryptococcosis granuloma and lung adenocarcinoma can appear similar on noncontrast chest CT. This study evaluates the use of an integrated model that was developed based on radiomic features combined with demographic and radiological features to differentiate pulmonary cryptococcosis nodules from lung adenocarcinomas.

Methods: Preoperative chest CT images for 215 patients with solid pulmonary nodules with histopathologically confirmed lung adenocarcinoma and cryptococcosis infection were collected from two clinical centers (108 cases in the training set and 107 cases in the test set divided by the different hospitals). Radiomics models were constructed based on nodular lesion volume (LV), 5-mm extended lesion volume (ELV), and perilesion volume (PLV). A demoradiological model was constructed using logistic regression based on demographic information (age, sex) and 12 radiological features (location, number, shape and specific imaging signs). Both models were used to build an integrated model, the performance of which was assessed using the test set. A junior and a senior radiologist evaluated the nodules. Receiver operating characteristic (ROC) curve analysis was conducted, and areas under the curve (AUCs), sensitivity (SEN), and specificity (SPE) of the models were calculated and compared.

Results: Among the radiomics models, AUCs of the LV, ELV, and PLV were 0.558, 0.757, and 0.470, respectively. Age, lesion number, and lobular sign were identified as independent discriminative features

---

^ ORCID: 0000-0002-8561-1729.

providing an AUC of 0.77 in the demoradiological model (SEN 0.815, SPE 0.642). The integrated model achieved the highest AUC of 0.801 (SEN 0.759, SPE 0.755), which was significantly higher than that obtained by a junior radiologist (AUC =0.689, P=0.024) but showed no significant difference from that of the senior radiologist (AUC =0.784, P=0.388).

Conclusions: An integrated model with radiomics and demoradiological features improves discrimination of cryptococcosis granulomas from solid adenocarcinomas on noncontrast CT. This model may be an effective strategy for machine complementation to discrimination by radiologists, and whole-lung automated recognition methods might dominate in the future.

Keywords: Radiomics; computed tomography (CT); pulmonary cryptococcosis granuloma; lung adenocarcinoma

Introduction

Pulmonary cryptococcosis is a major opportunistic invasive mycosis caused predominantly by *Cryptococcus neoformans* or *Cryptococcus gattii* (1). While immunocompromised individuals are susceptible to pulmonary cryptococcal infections, approximately 30% of the population with normal immunity may not exhibit any significant symptoms or laboratory findings. Therefore, chest computed tomography (CT) serves as a critical tool in the diagnosis and evaluation of pulmonary cryptococcosis (2-5).

Evaluating patients with CT Chest for cryptococcosis can be challenging since these patients are commonly misdiagnosed with lung cancer due to similar radiological manifestations, such as pulmonary nodules, cavities, and enlarged lymph nodes (2-5). In 57–82% of pulmonary cryptococcosis cases, the disease manifests as a nodule or mass with necrosis and cavities, lobulation, and air bronchogram signs on CT (3,6), mimicking peripheral lung cancer. Unexperienced radiologists are more likely to misdiagnose. As a result, some patients undergo unnecessary surgeries with the risk of disseminated cryptococcal infection. A study of 18 fluorine-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET)-CT revealed that approximately 88% of lung lesions in pulmonary cryptococcosis demonstrate high FDG uptake, mimicking a possible malignant condition (7), further limiting the use of PET in the diagnosis. Moreover, the treatment for patients with cryptococcal pneumonia requires antifungal therapy while prompt resection in patients with early-stage carcinoma usually improves prognosis. Therefore, it can be very helpful to differentiate the two diseases accurately with an effective non-invasive method.

Recently, radiomics has been gaining traction as it provides a quantitative method to identify the intrinsic heterogeneity of a lesion and ensures precise classification (8-10). Several previous studies have revealed radiomics analysis with the potential to differentiate between benign and malignant pulmonary nodules on noncontrast CT (11-16), including infectious diseases and carcinoma. For example, Beig *et al.* proposed a model combining radiomics features of intranodular and perinodular regions on lung CT images, and the accuracy of distinguishing adenocarcinomas from granulomas improved to 0.80 compared with the accuracy of 0.6 from two radiologists (17). However, only a few studies have focused on radiomics or computerized feature-based classification of fungal infections and carcinomas (16,18-20). Moreover, small sample sizes and single-center designs have rendered the results unconvincing. In this study, a multi-institution design with


Highlight box

Key findings
- An integrated radiomics model could distinguish pulmonary cryptococcosis from adenocarcinoma.

What is known and what is new?
- Radiomics features has the potential to differentiate between benign and malignant pulmonary nodules.
- Radiomics combined with demoradiological features could improve the diagnostic performance in the differentiation of disease mimickers, with a multi-institution design.

What is the implication, and what should change now?
- This integrated model has the potential to help make better clinical decision, and need to be further validated and widely applied.

1792

Hu et al. Radiomics model for pulmonary cryptococcosis *vs.* lung cancer
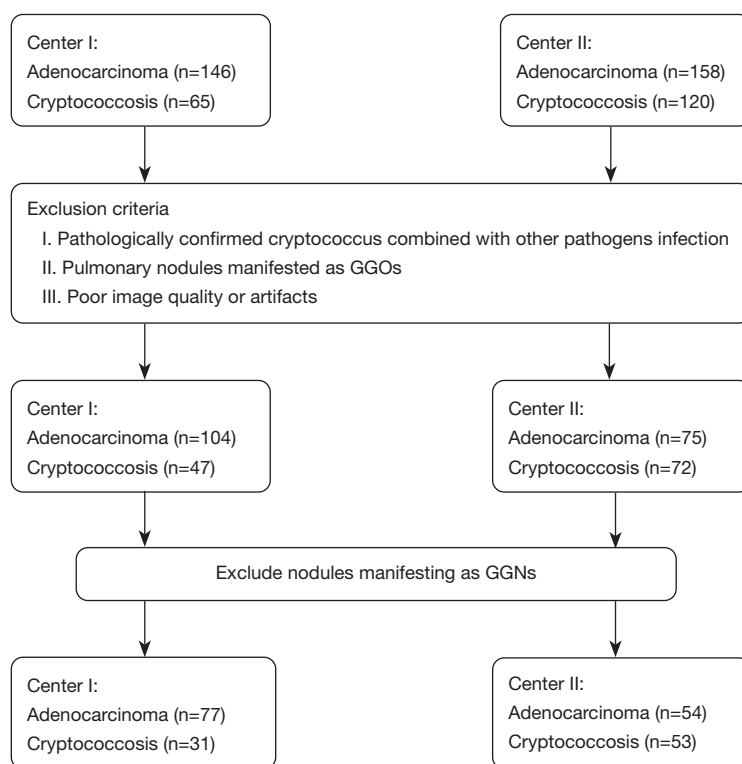
Figure 1 Flowchart of the case selection. GGO, ground-glass opacity; GGN, ground-glass nodule.

larger sample sizes were used to investigate the ability of a CT-based radiomics classifier to differentiate pulmonary cryptococcosis nodules and lung adenocarcinomas. We present this article in accordance with the TRIPOD reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-389/rc).

Methods

Study population and image acquisition

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics committees of Shanghai Chest Hospital [No. KS(Y)1990] and the other clinical center was informed and agreed with this study. Individual consent for this retrospective analysis was waived. Noncontrast lung CT images were collected from two clinical centers (Huashan Hospital, Center I; Shanghai Chest Hospital, Center II) between 2015 and 2019, the imaging parameters are listed in Table S1.

Non-immunocompromised patients with pulmonary cryptococcosis confirmed by histopathologic examination

via surgical resection or biopsy were screened using the electronic medical record, and cases with lung adenocarcinoma were randomly selected and matched with the control group at a ratio of 1:3. The inclusion criteria were as follows: available baseline preoperative noncontrast CT scans demonstrating solid pulmonary nodules; pulmonary nodules <4 cm in diameter; and a reconstructed slice thickness of CT scans <1.5 mm. The exclusion criteria were as follows: pathologically confirmed cryptococcus combined with other pathogen infections; pulmonary nodules manifested as ground-glass opacities (GGOs); and poor image quality or artifacts.

Data for 215 patients were collected from the two centers. The data for 108 patients from Center I constituted the training set for the model (31 pulmonary cryptococcoses and 77 lung cancer). Data for 107 patients from Center II were used as an independent test set to assess model performance (53 pulmonary cryptococcoses and 54 lung adenocarcinomas) (*Figure 1*).

CT image annotation and feature extraction

Three types of volumes of interest (VOIs) were generated,
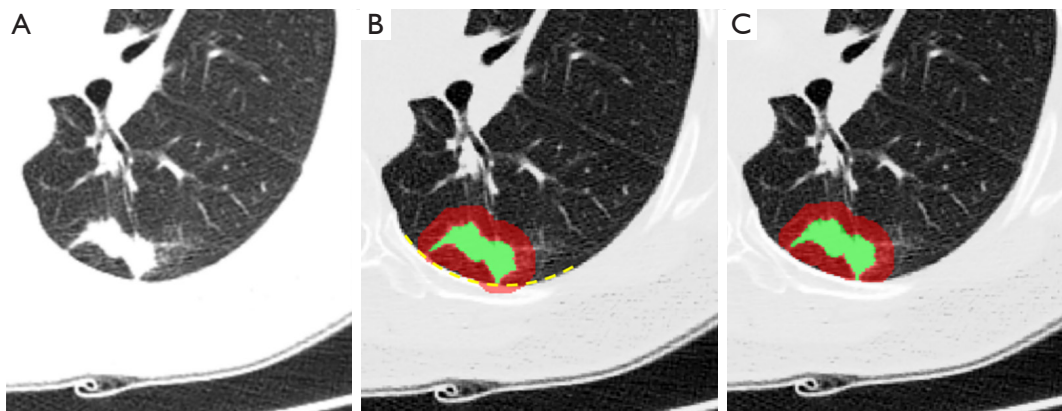
Figure 2 Illustration of three types of the VOI. (A) An irregular nodule appears on the CT image (lung window). (B) LV (green region) was obtained by manually drawing the whole lesion. The boundary of the LV was increased by 5 mm, and the ELV (red and green regions) was generated after excluding the regions outside the lung (yellow line). (C) PLV (red region) was obtained by subtracting LV (green region) from ELV. VOI, volume of interest; LV, lesion volume; ELV, extended lesion volume; PLV, perilesion volume.

as indicated in *Figure 2*. Lesions were manually delineated by a radiologist (Hu B, with 7 years of experience) using three-dimensional (3D) slicer software (version 4.8.1), and each lesion volume (LV) of interest was obtained. Then, the extended lesion volume (ELV) of interest was generated using the dilation operation to extend the boundary of the LV by 5 mm. Regions of tissues outside the lung were excluded from the ELV by manual correction. In addition, the perilesion volume (PLV) of interest was obtained by subtracting the LV from the ELV.

Radiomics features were extracted from the LV, ELV, and PLV. For each type of VOI, 851 radiomics features were extracted and classified as follows (21): (I) 14 shape features describe the geometric information of the VOI; (II) 18 statistical features represent the characteristics of the intensity histogram within the VOI; (III) 75 texture features quantify the spatial heterogeneity of the intensity within the VOI; and (IV) 744 wavelet features derived from the wavelet filtered images highlight the local details or approximate shape of the images. The radiomics features were extracted using the open-source tool Pyradiomics (version 2.1.2), and the details can be found at the public website (https://pyradiomics.readthedocs.io/en/2.1.2/features.html).

Radiomics model building

Radiomics models were built for LV, ELV, and PLV (denoted as M_{Rads_LV}, M_{Rads_ELV}, and M_{Rads_PLV}, respectively); the procedure included two steps: feature selection and model training.

For feature selection, Spearman's correlation coefficient (SCC) of each radiomics feature pair was calculated. The radiomics feature pair with SCC >0.9 was considered to be highly correlated (22). In the feature pair, the feature with a high mean SCC was considered redundant and therefore was eliminated. To decrease the risk of overfitting, the minimum redundancy maximum relevance (mRMR) feature ranking algorithm was used to score each feature according to its importance in discriminating the two diseases (23). Only the top-ranking features were selected for model building, and the number of features was limited to 1/10th of the number of cases (24).

Selected radiomics features were input in the least absolute shrinkage and selection operator (LASSO) for radiomics model building (25). L1 regularization in logistic regression facilitated LASSO-mediated compression of the selected features, thereby improving the performance of the radiomics model. The hyperparameter λ controlled the degree of feature compression. For each type of VOI, a radiomics model was trained by LASSO, and the hyperparameter λ was determined using fivefold cross-validation in the training set. The radiomics model was built as the linear weighted sum of the radiomics features.

Demoradiological and integrated models

Demographic features, including age and sex, were collected. Radiological features, including the location, number and shape of lesions and whether the lesions

Table 1 Demographic information of two groups

Variables	Training set			Test set		
	Pulmonary cryptococcosis (n=31)	Lung cancer (n=77)	P value	Pulmonary cryptococcosis (n=53)	Lung cancer (n=54)	P value
Age (years), mean ± SD	46.42±12.52	59.74±9.15	<0.001	53.60±10.58	60.87±7.94	<0.001
Sex			0.490			0.485
Male	18	39		20	24	
Female	13	38		33	30	

SD, standard deviation.

had the spicular sign, lobular sign, air bronchus sign, pseudocavitation, vascular convergence sign, halo sign, calcification, satellite lesions, or pleural change were documented by two qualified radiologists (Li Y and Zhu L, each with 20 years of experience) blinded to the diagnosis. Univariate logistic regression was used to assess the disease-discriminative ability of each demographic and radiological feature. The significant features identified by univariate logistic regression were further input into multivariate logistic regression to identify independent discriminative features and build a demoradiological model, $M_{demoradiological}$.

The integrated model $M_{integrated}$ was built by combining the predictions of the radiomics model and the independent discriminative demoradiological features using multivariate logistic regression.

Model testing and statistical analysis

Performance of the models was assessed in the independent test set. Moreover, two radiologists (Tang Y and Li Y, with 10 and 20 years of experience, respectively) blinded to the histological results were invited to evaluate the pulmonary nodules on a scale of 1–4, indicating benign to malignant. The test performances of the models and radiologists were quantified by the area under the receiver operating characteristic (ROC) curve (AUC), accuracy (ACC), sensitivity (SEN) and specificity (SPE). Differences

between the ROC curves were assessed using the DeLong test (26). The Mann-Whitney U test was used for numerical characteristics in statistical tests of demographics, and the Fisher's exact test was applied for categorical characteristics. P<0.05 was considered statistically significant. The feature selection, model building, model performance test, and statistical analysis were conducted using R (v. 3.6.2, https://www.r-project.org/).

Results

Demographic information of subjects

The patients' profiles are listed in *Table 1*. The mean age of the patients with pulmonary cryptococcosis was significantly younger than that of the patients with lung adenocarcinoma in both the training and test sets. The sex distribution did not show a significant difference between the two types of diseases in both sets.

Diagnostic performance of the radiomics model

A total of 851 radiomics features were extracted from each kind of VOI. Among these, 4, 8, and 11 radiomics features were selected and trained for the LV-, ELV-, and PLV-based radiomics models, respectively. The formula of the radiomics models was as follows:

$$
\begin{aligned}
M_{Rads\_LV} = &\ 1.24 + 1.93 \times original_{shape_{Elongation}} - 4.77 \times wavelet.LHH_{glcm_{MCC}} \\
&+ 6.87 \times wavelet.HHH_{firstorder_{Skewness}} \\
&+ 0.0973 \times wavelet.LLL\_firstorder\_Kurtosis
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
M_{Rads\_ELV} =\ & -0.118 + 4.11 \times original\_shape\_Elongation \\
& + 4.66 \times wavelet.LLH\_glcm\_Correlation \\
& - 2.02 \times wavelet.LHL\_glcm\_MCC + 8.04 \times wavelet.LHH\_glcm\_Idn \\
& + 0.528 \times wavelet.HLL\_firstorder\_Skewness \\
& - 2.99 \times wavelet.HLH\_firstorder\_Mean \\
& + 1.049 \times wavelet.HHH\_glszm\_SmallAreaEmphasis + 1.362e \\
& - 4 \times wavelet.LLL\_glcm\_ClusterTendency
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
M_{Rads\_PLV} =\ & -125 + 9.46 \times original.shape.Elongation \\
& + 9.85 \times wavelet.LLH.glcm.Idmn \\
& - 0.032 \times wavelet.LHL.firstorder.Mean \\
& + 24.1 \times wavelet.LHL.glcm.Idn \\
& + 0.554 \times wavelet.LHH.firstorder.Mean \\
& + 17.5 \times wavelet.HLL.glcm.Imc1 \\
& - 3.96 \times wavelet.HLH.firstorder.Mean \\
& - 8.02 \times wavelet.HHL.firstorder.Skewness \\
& - 1.22 \times wavelet.HHH.firstorder.Mean + 2.516e \\
& - 4 \times wavelet.LLL.firstorder.RootMeanSquared + 7.110e \\
& - 7 \times wavelet.LLL.glcm.ClusterShade
\end{aligned}
\tag{3}
$$

The AUCs of the LV-based radiomics model (M_{Rads_LV}) and the PLV-based radiomics model (M_{Rads_PLV}) were 0.558 and 0.470, respectively. Compared to the former, the ELV-based radiomics model (M_{Rads_ELV}) achieved a higher AUC of 0.757.

Diagnostic performance of the demoradiological model

The details of the radiological features that appeared in each case were documented and summarized for both the training and test sets (Table S2). Univariate logistic regression demonstrated that the significant demoradiological features predicting disease classification were age, number of lesions, spicular sign, lobular sign, and satellite lesions. Results of multivariate logistic regression identified that age, number, and lobular sign were independent discriminative features. The diagnostic performance of the demoradiological model is shown in *Table 2*. The AUC of $M_{demoradiological}$ was 0.770, and the formula of $M_{demoradiological}$ was:

$$
\begin{aligned}
M_{demoradiological} =\ & 0.151 \times Age - 2.54 \times Number \\
& + 2.06 \times Lobular\ sign - 7.564
\end{aligned}
\tag{4}
$$

Diagnostic performance of the integrated model

Combining the predictions of the radiomics model and the independent discriminative demoradiological features, an integrated model ($M_{integrated}$) was constructed. It achieved an AUC of 0.801, which was the highest AUC among all models. The formula of $M_{integrated}$ was:

$$
\begin{aligned}
M_{integrated} =\ & 0.144 \times Age - 2.33 \times Number + 1.77 \times lobular\ sign \\
& + 0.920 \times Prediction_{Rads\_ELV} - 7.62
\end{aligned}
\tag{5}
$$

$Prediction_{Rads\_ELV}$ is the prediction of M_{Rads_ELV}.

Comparison of diagnostic ability between classifier models and radiologists

In terms of the diagnostic performance of radiologists, AUCs were 0.784 and 0.689 for the senior and junior radiologists, respectively. The AUC of M_{Rads_ELV} was compared to the AUC of $M_{demo-radiological}$ (P=0.579), radiologist 1 (P=0.165), and radiologist 2 (P=0.654). The AUC of $M_{integrated}$ was significantly higher than that of radiologist 1 (P=0.024) and $M_{demo-radiological}$ (P=0.015) but had not significantly different

Table 2 Univariate and multivariate logistic regression results of demoradiological features

Feature	Univariate		Multivariate	
	OR (95% CI)	P value	OR (95% CI)	P value
Age	1.14 (1.07–1.21)	<0.001	1.15 (1.07–1.23)	<0.001
Sex	0.74 (0.32–1.72)	0.486	–	–
Upper lobe	–	–	–	–
Middle lobe	0.92 (0.17–5.05)	0.920	–	–
Lower lobe	0.41 (0.17–0.99)	0.050	–	–
Number of lesions	0.09 (0.03–0.3)	<0.001	0.07 (0.01–0.38)	0.002
Shape (irregular, regular)	1.15 (0.5–2.68)	0.739	–	–
Spicular sign	3.09 (1.26–7.58)	0.014	1.37 (0.39–4.82)	0.628
Lobular sign	3.63 (1.48–8.92)	0.005	7.98 (1.99–32.01)	0.003
Air bronchus sign	1.67 (0.6–4.62)	0.326	–	–
Pseudocavitation	1.36 (0.45–4.11)	0.582	–	–
Calcification	–	0.992	–	–
Halo sign	0.53 (0.19–1.46)	0.220	–	–
Satellite lesions	0.19 (0.05–0.7)	0.013	0.21 (0.04–1.16)	0.073
Vascular convergence sign	4.11 (0.89–18.99)	0.070	–	–
Pleural change	–	–	–	–
Indentation	1.65 (0.64–4.22)	0.296	–	–
Attachment	1.32 (0.37–4.73)	0.670	–	–

The OR of location was calculated by considering the upper lobe as a reference. The OR of calcification could not be calculated, as only one patient in each dataset (with lung adenocarcinoma) had calcification. OR, odds ratio; CI, confidence interval.

from that of radiologist 2 (P=0.388).

The AUC, accuracy, sensitivity, and specificity of each model and two radiologists are listed in *Table 3*. ROC curves are illustrated in *Figure 3*.

Among 107 cases in the external test set, 20 were correctly classified by the integrated model but misdiagnosed by at least one radiologist. Sixteen cases were classified incorrectly by the model but correctly diagnosed by at least one radiologist. Six cases of pulmonary cryptococcosis showed irregular shape, spiculation sign, air bronchi sign, and vascular convergence sign but were neither accurately predicted by the classifier model nor the radiologists. The representative easy-mimicking pulmonary nodules are shown in *Figure 4*.

Discussion

In clinical practice, distinguishing the solid pulmonary

nodules of cryptococcosis from carcinomas on CT is challenging. In this radiomics study with 215 cases, we constructed an integrated model combining radiomics, demographic, and radiological features, which achieved an AUC of 80.1% in the independent test set. The performance of the integrated model surpassed the level of the senior radiologist or the model based on traditional radiological features.

Lung squamous cell carcinoma typically presents as a central lung cancer and is often accompanied by necrosis or cavitation. These imaging features are different from those of pulmonary cryptococcosis. According to the location and characteristics of the disease, a large overlap between pulmonary adenocarcinoma and cryptococcosis manifested as solid nodules can be observed. Therefore, we enrolled only adenocarcinoma cases in the present study. For radiologists, recognizing radiological features is critical in differentiating between pulmonary cryptococcosis

Table 3 Performance of models and radiologists

Model/radiologist	AUC	ACC	SEN	SPE
M_{Rads_LV}	0.558	0.383	0.296	0.472
M_{Rads_ELV}	0.757	0.757	0.704	0.811
M_{Rads_PLV}	0.470	0.523	0.167	0.887
$M_{demoradiological}$	0.770	0.729	0.815	0.642
$M_{integrated}$	0.801	0.757	0.759	0.755
Radiologist 1	0.689	0.654	0.889	0.415
Radiologist 2	0.784	0.720	0.852	0.585

AUC, area under the receiver operating characteristic curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; LV, lesion volume; ELV, extended lesion volume; PLV, perilesion volume.
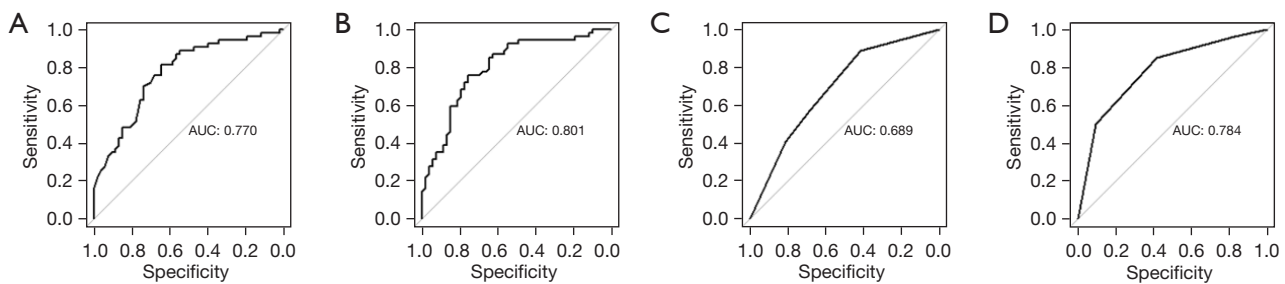

Figure 3 ROC curves of models and radiologists. (A) $M_{demoradiological}$, (B) $M_{integrated}$, (C) Radiologist 1, (D) Radiologist 2. ROC, receiver operating characteristic; AUC, area under the ROC curve.

and lung cancer. Liu *et al.* demonstrated that clinically relevant radiological features, including attachment to the vessel, nodule location, border definition, and concavity, could effectively determine the lung cancer risk of solitary pulmonary nodules (27). Several studies have also revealed typical signs of pulmonary cryptococcosis, which can be noted as pulmonary nodules with segmental or lobar consolidation or a reticulonodular pattern of opacities (4,6,28,29). However, there is some overlap between the two diseases. Compared to lung cancer, nodules with cryptococcosis are more common in the outer field of the lung, with lower tension, rare deep lobulation, and more air bronchogram signs (5). In the current study, radiological features, such as lobular signs, satellite lesions, and vascular convergence signs, were shown to be significantly different between the two groups after logistic regression, consistent with previous studies (28,30-32). Moreover, when the age factor was considered, the AUC improved to 0.770, which was even better than the performance of a junior radiologist. These results suggest the importance of precise

identification of radiological features in diagnosis which may be challenging for radiologists with limited experience.

Therefore, with a growing need for practical and automated means, radiomics has become a solution. Lesion segmentation is a principal procedure in construction of a radiomics classifier model. In this study, we defined three types of VOIs (LV, ELV, and PLV) following the method of Beig *et al.* (17); the ELV-based radiomics model showed better performance than the others, indicating that the perinodular region could possess valuable information for improving the efficiency of intranodular radiomics analysis. Among radiomics studies related to chest CT, the majority have focused solely on lung nodule shape and texture analysis (11,33,34). Nonetheless, information in the nodular-surrounding tissues, including characteristics of tumor-infiltrating lymphocytes and tumor-associated stromal macrophages not reflected in the lesion, is yet to be determined (11). Thus, assessment of the entire tumor together with the surrounding tissue might provide complementary
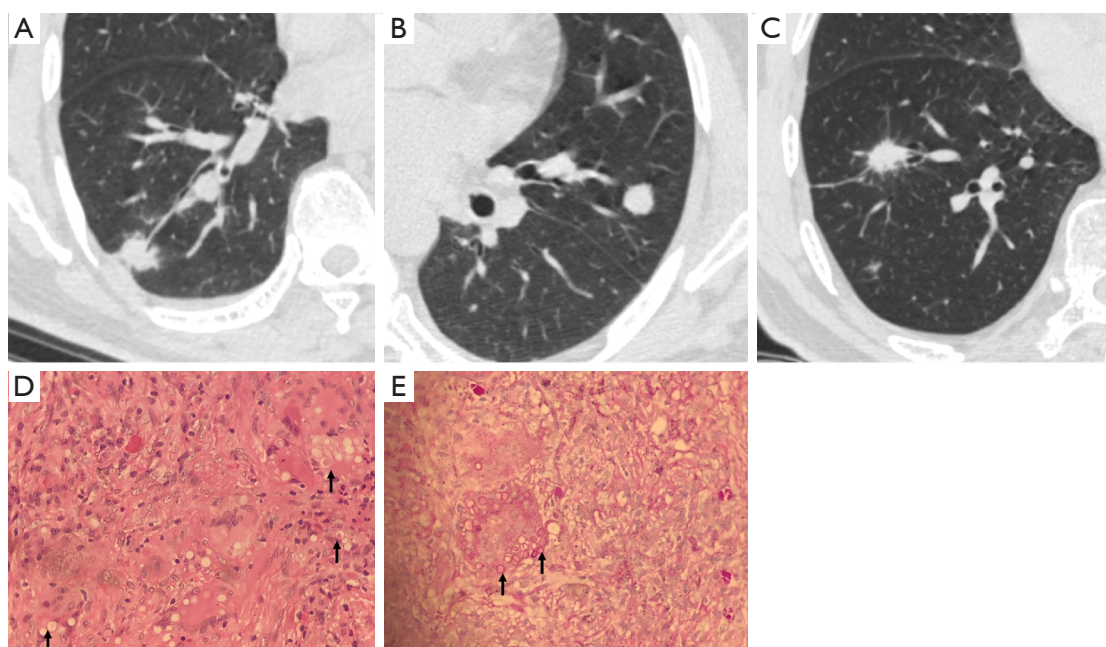
Figure 4 Representative easy-mimicking pulmonary nodules on CT images. (A) A 59-year-old male patient with pulmonary cryptococcosis was misdiagnosed with a malignant mass by both radiologists but was correctly identified by the classifier model. (B) A 60-year-old male patient with lung cancer was misdiagnosed with a benign nodule by both radiologists but was correctly recognized by the classifier model. (C) A 64-year-old male patient misdiagnosed with pulmonary carcinoma by both radiologists and the integrated model was pathologically confirmed to have cryptococcal granuloma. (D) Histopathological examination (hematoxylin erythrosine staining) of lung tissue reveals granulomatous inflammation. The fungal cells are round, lightly basophilic, fill the alveolar space and are surrounded by a clear space that represents the capsular material (see arrows, ×200). (E) Positive periodic acid-Schiff staining of the yeast wall, suggesting pulmonary cryptococcosis (see arrows, ×200).

information, which is crucial for accurate diagnosis.

In addition to VOI selection, coalescing of information is a strategy for improving classifier performance. In this study, discrimination based on integrated models combined with demographic, radiological, and radiomics features was superior to any other models, indicating that human-machine integration is important for decision-making. Additionally, it is challenging to identify subtle differences in image features by visualization, even for senior radiologists, and radiomics signatures based on quantitative parameters can be helpful. Moreover, radiological features such as lesion number and location provide valuable information at a "global" level; some unique manifestations such as "satellite lesions" are also key clues in the diagnosis. On the other hand, it is challenging for machines alone to conduct precise identification between two diseases, rendering the experience of expert radiologists critical with respect to traditional qualitative features and in interpreting low-quality imaging data (35).

The present study has several limitations. First, we only considered age and sex as demographic features. If other risk factors, such as smoking history and epidemiological history, are considered, the diagnostic performance of the model could be improved. Second, the case cohort is not generalizable, although a large number of cases were collected in two clinical centers, the sample size of pulmonary cryptococcus infection was small due to its low incidence rate, and a further external validation is needed. Third, although the integrated model provided a combination of experts' experiences and imaging features, it is still a semiautomatic method. In terms of feasibility and practicality, whole-lung automated recognition methods might dominate in the future.

Conclusions

In conclusion, radiomics features of pulmonary nodules can distinguish cryptococcosis granulomas from solid

adenocarcinomas on noncontrast CT and improve classification accuracy. Additionally, incorporating perinodular texture features with intranodular texture improves the diagnostic performance of our novel classifier. Importantly, machine complementation of humans could be an effective strategy in the differentiation of disease mimickers.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-389/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-389/dss

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-389/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-23-389/coif). AW has done some consulting work for Noah Medical, Ambu, and Medtronic, and receives payment or honoraria for speaker bureaus for Biodesix, and also done some medicolegal work in the area of lung nodules. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics committees of Shanghai Chest Hospital [No. KS(Y)1990] and the other clinical center was informed and agreed with this study. Individual consent for this retrospective analysis was waived.

References

1. Setianingrum F, Rautemaa-Richardson R, Denning DW. Pulmonary cryptococcosis: A review of pathobiology and clinical aspects. Med Mycol 2019;57:133-50.
2. Deng H, Zhang J, Li J, et al. Clinical features and radiological characteristics of pulmonary cryptococcosis. J Int Med Res 2018;46:2687-95.
3. Qu Y, Liu G, Ghimire P, et al. Primary pulmonary cryptococcosis: evaluation of CT characteristics in 26 immunocompetent Chinese patients. Acta Radiol 2012;53:668-74.
4. Fox DL, Müller NL. Pulmonary cryptococcosis in immunocompetent patients: CT findings in 12 patients. AJR Am J Roentgenol 2005;185:622-6.
5. Zhang Y, Li N, Zhang Y, et al. Clinical analysis of 76 patients pathologically diagnosed with pulmonary cryptococcosis. Eur Respir J 2012;40:1191-200.
6. Xie LX, Chen YS, Liu SY, et al. Pulmonary cryptococcosis: comparison of CT findings in immunocompetent and immunocompromised patients. Acta Radiol 2015;56:447-53.
7. Wang SY, Chen G, Luo DL, et al. (18)F-FDG PET/CT and contrast-enhanced CT findings of pulmonary cryptococcosis. Eur J Radiol 2017;89:140-8.
8. Lee SH, Lee SM, Goo JM, et al. Usefulness of texture analysis in differentiating transient from persistent part-solid nodules (PSNs): a retrospective study. PLoS One 2014;9:e85167.
9. Chae HD, Park CM, Park SJ, et al. Computerized texture analysis of persistent part-solid ground-glass

nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. Radiology 2014;273:285-93.

10. Fan L, Fang M, Li Z, et al. Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. Eur Radiol 2019;29:889-97.

11. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114:345-50.

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

13. Wu L, Gao C, Xiang P, et al. 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.

14. Uthoff J, Nagpal P, Sanchez R, et al. Differentiation of non-small cell lung cancer and histoplasmosis pulmonary nodules: insights from radiomics model performance compared with clinician observers. Transl Lung Cancer Res 2019;8:979-88.

15. Khorrami M, Bera K, Thawani R, et al. Distinguishing granulomas from adenocarcinomas by integrating stable and discriminating radiomic features on non-contrast computed tomography scans. Eur J Cancer 2021;148:146-58.

16. 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. Cancer Imaging 2020;20:45.

17. Beig N, Khorrami M, Alilou M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. Radiology 2019;290:783-92.

18. 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. Cancers (Basel) 2021;13:2781.

19. Zhang Y, Chu Z, Yu J, et al. Computed tomography-based radiomics for identifying pulmonary cryptococcosis mimicking lung cancer. Med Phys 2022;49:5943-52.

20. Zhang ZX, Mu XY, Yu J, et al. Establishment and evaluation of a CT-based radiomic model for AIDS-associated pulmonary cryptococcosis. BMC Med Imaging 2022;22:185.

21. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:e104-7.

22. Xia W, Hu B, Li H, et al. Multiparametric-MRI-Based Radiomics Model for Differentiating Primary Central Nervous System Lymphoma From Glioblastoma: Development and Cross-Vendor Validation. J Magn Reson Imaging 2021;53:242-50.

23. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27:1226-38.

24. Nie K, Shi L, Chen Q, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI. Clin Cancer Res 2016;22:5256-64.

25. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010;33:1-22.

26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.

27. Liu Y, Wang H, Li Q, et al. Radiologic Features of Small Pulmonary Nodules and Lung Cancer Risk in the National Lung Screening Trial: A Nested Case-Control Study. Radiology 2018;286:298-306.

28. Soubani AO. The evaluation and management of the solitary pulmonary nodule. Postgrad Med J 2008;84:459-66.

29. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology 2017;284:228-43.

30. Wu W, Pierce LA, Zhang Y, et al. Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: a case-control study. Eur Radiol 2019;29:6100-8.

31. Edey AJ, Hansell DM. Incidentally detected small pulmonary nodules on CT. Clin Radiol 2009;64:872-84.

32. Gould MK, Ananth L, Barnett PG, et al. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest 2007;131:383-8.

33. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology 2016;281:947-57.

34. He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 2016;6:34921.

35. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. Eur J Radiol 2017;86:297-307.

Table S1 Imaging parameters of CT scans

Variables	Shanghai Chest Hospital				Shanghai Huashan Hospital			
	Philips Brilliance 64	GE Discovery CT750 HD	Philips Ingenuity CT	Philips iCT 256	Siemens Sensation 64	GE Discovery CT	UIH uCT 760	Philips iCT 256
Tube voltage (kVp)	120	120	120	120	120	120	120	120
Tube current (mA)	Auto	Auto	Auto	Auto	Auto	Auto	Auto	Auto
Slice Thickness	1	0.625/1.25	1	1	1.5	1.25	1.5	1.5
Matrix	1024×1024	512×512	1024×1024	1024×1024	512×512	512×512	512×512	512×512
Convolution Kernel	Very sharp	Very sharp	Very sharp	Very sharp	Very sharp	Very sharp	Very sharp	Very sharp

Table S2 Number of cases with radiological features in the training and independent test sets

Feature	Training set		Test set	
	Cryptococcus (n=31)	Adenocarcinoma (n=77)	Cryptococcus (n=53)	Adenocarcinoma (n=54)
Location				
Upper lobe	11	42	19	26
Middle lobe	2	7	3	3
Lower lobe	18	28	31	25
Number				
Single	19	73	43	29
Multiple	12	4	10	25
Shape				
Irregular	13	35	20	33
Regular	18	42	33	21
Spicular sign	9	43	18	24
Lobular sign	9	46	19	40
Air bronchus sign	6	22	6	24
Pseudocavitation	5	16	1	6
Calcification	0	1	0	1
Halo sign	8	12	13	10
Satellite lesions	7	4	6	1
Vascular convergence sign	2	17	5	20
Pleural change				
Indentation	15	45	18	32
Attachment	5	12	2	20