

Background

The cryptococcal granulomas are very similar as the small malignant pulmonary nodules in morphology and pulmonary position. The high similarity can make the early diagnosis of lung cancer challenging, and many patients with cryptococcal granulomas suffer the pain of biopsy confirmation. A deep learning method for distinguishing two types of nodules was established to address the challenging, and facilitate the fast, accurate diagnosis in clinic.

Study Design and Methods

We conducted a single-center study including 714 contrast CT scans from 295 patients (87 with cryptococcosis and 208 with lung cancer). The disease status of each patient was biopsy confirmed. A dual-path deep learning (DL) pipeline mimicking the diagnosis process of the radiologists was built to differentiate between the cryptococcal granulomas and the small malignant nodules based on nodule and lung area from the contrast CT scans. The single-path DL pipeline as well as the radiomics pipeline was established respectively to explore the diagnosis performance solely based on the nodule area. The transfer learning training strategy was employed to train both DL pipelines for achieving the better results.

Results

All the patients were randomly divided into the train dataset and the test dataset with the ratio 4:1. Three classification pipelines were evaluated by 5-fold cross validation on the train dataset, then the best model was selected to do the performance evaluation on the test dataset. The mean area under the receiver operating characteristic curve (AUC) was the performance metric. Compared to the single-path DL pipeline (mean validation AUC 0.85 ± 0.02 , test AUC 0.80) and the radiomics pipeline (mean validation AUC 0.79 ± 0.01 , test AUC 0.75), the dual-path DL pipeline achieved the best mean validation AUC 0.88 ± 0.01 , and the best test AUC 0.83. The radiomics pipeline can achieve competitive results as the single-path DL pipeline, and 12 important features were selected for the further interpretation.

Interpretation

The dual-path DL pipeline mimicking the diagnosis process in clinic provided a relative quick, reliable approach of distinguishing cryptococcal granulomas from small malignant pulmonary nodules based on the lung area and the nodule area.

Distinguish Solitary Cryptococcosis from Malignant Pulmonary Nodules based on Lung Contrast CT Using Deep Learning

Xinyu Guo (1) *
Youbing Yin (1)

xinyug@keyamedna.com
yin@keyamedna.com

(1) The Department of Research & Development, Keya Medical, Seattle, USA

* indicates the corresponding author

Key Words

Cryptococcosis, pulmonary nodules, deep learning, computed tomograph

Abbreviation

CT = computed tomography, DL = deep learning,

Lung cancer remains the leading cause of cancer, with high incidence and mortality, all over the world [1]. Over half of the patients die within one year of diagnosis and the 5-year survival is less than 18% [2]. Early diagnosis of malignant lung nodules is critical for the treatment. CT imaging is an effective non-invasive diagnosis tool and is widely used for lung cancer screening. However, though several biomarkers such as nodule sizes might be associated with malignance, it is still challenging to differentiate malignant nodules from some other benign nodules which share similar CT manifestations. Cryptococcosis, which is an infection caused by yeast like fungus, is one common false positive during the lung cancer screen [3–8]. One subtype of cryptococcosis may present a single large nodule on CT, leading to a high misdiagnose rate into lung cancer [4]. The main manifestations of CT findings associated with cryptococcosis were mainly classified into three patterns, namely nodular, pneumonia and mixed morphological characteristics [26]. Radiologists not only need to review the nodule itself and but also look for the other associated findings in whole lung regions in order to make more certain decision[27][28].

Radiomics [9–17] or deep learning [18–22] techniques have been widely adopted to predict the malignance of lung nodules from images due to the powerful feature extraction and classification capability. For example, in [24] the authors used the intranodular and perinodular radiomic features was constructed to classify adenocarcinomas from granuloma. The performance was evaluated using the area under the receiver operating characteristic (AUC) and it was 0.8 by combining radiomics of intranodular with perinodular regions. It should be noted that the authors also tried a

simple DL model for the classification to achieved comparable performance with the radiomics model using intramodular features only. However, the parameters for the perinodular areas had to pre-defined and the top radiomics features needed to be pre-selected to achieve satisfactory performance. It might be difficult to define suitable perinodular areas if the associated findings may be observed in other regions not adjacent to nodules. In [23], the authors adopted a complex but end-to-end DL model in order to distinguish malignant lung nodules from benign ones and achieved the satisfactory performance. However, the appearance of both types of nodules were quite different in the dataset, making the classification relatively easy.

In this study, we aimed to differentiate solitary benign cryptococcal nodules from malignant lung cancer nodules based on chest CT. Both radiomics and DL methods were adopted for the task and their performance was thoroughly compared. Specifically, a novel deep learning framework was proposed to address several challenging issues in this task. First, the framework takes the inputs of both nodule region and whole lung images to take account of both the local and global information considering that CT findings associated with cryptococcosis may appear in other parts of lungs in addition to the nodule neighboring regions. Second, our deep learning framework used a new loss function to directly optimize AUC to achieve high performance

Materials and Methods

Patients

The current retrospective study included patients who were pathologically confirmed with either lung cancer or cryptococcosis. Patients with the following criteria were first used for selection in the database: availability of pathologic report through thoroscopic wedge resection, availability of a diagnostic thoracic contrast enhanced CT image with the slices thickness ≤ 2 mm, and presence of a solitary pulmonary nodule with equivalent nodule diameter (equivalent diameter assuming that a nodule is a sphere) between 15mm and 4 cm. 299 patients were selected from the database between June 2009 and August 2019. We then further visually checked the images to select the patients with similar morphological characteristics of nodules between the lung cancer and cryptococcosis groups. The morphological characteristics included lobulation, spiculation, pleural stretch, mixed ground-glass opacity nodules with partial solid component, intranodular necrosis, heterogeneous enhancement, local bronchial wall thickening, or nodules presenting in the corresponding hilum

of the lung. The final selected patients must include nodule with at least two of characteristics above. Four patients were excluded so that 295 patients were included in the final cohort. In order to increase the amount of data, for each patient, all contrast CT scans reconstructed with different convolutional kernels, but patients' position were the same across all scans were selected, this results in total 714 contrast CT scans in the whole dataset. Among 295 patients, 208 patients were diagnosed as lung cancer and 87 were diagnosed as cryptococcosis. All patients were divided into the training set as well as the test set with the ratio 4:1. The distribution of two types of patients were almost the same across two sets. As Figure 1 illustrated, there were 58 patients with 147 scans in the test set, 237 patients with 567 scans in the training set. The patients in the training set were divided into 5 folds evenly. The distribution of two types patients in both sets were presented with color bars. The green bar indicated the cryptococcosis patients while the orange bar indicated the lung cancer patients. The number below each orange bar indicated the number of contrast CT scans of the relative set/fold.

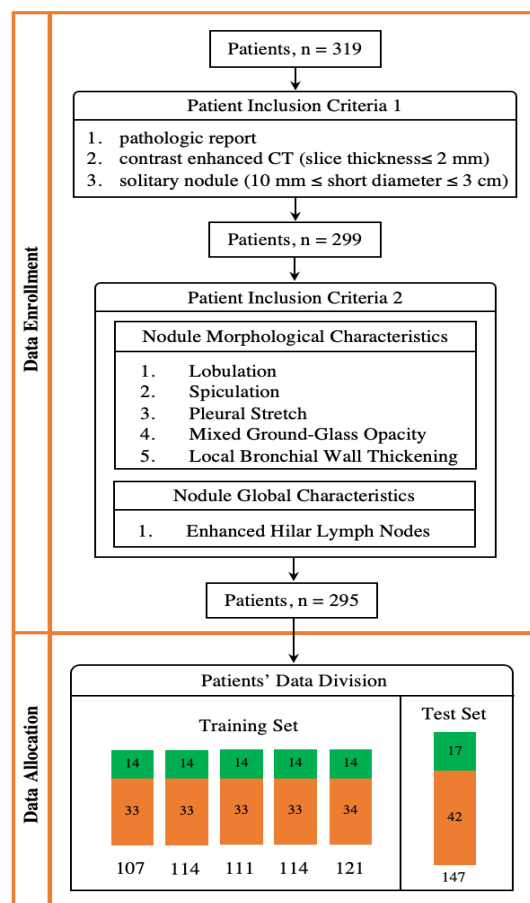


Figure 1. Flow diagram of data enrollment and data allocation

CT Examinations

CT examinations were performed at the supine position using the CT scanners from one of five CT manufactures with the following distributions: 96 patients from PHILIPS (Amsterdam, Netherlands) with 212 scans, 110 patients from TOSHIBA (Tokyo, Japan) with 286 scans, 16 from SIEMENS (Munich, Germany) with 36 scans, 72 from General Electric System (Boston, US) with 176 scans, and one patient from UNITED IMAGING HEALTHCARE (Shanghai, China) with four scans. The size of the reconstruction matrix was 512×512 . The tube voltage ranged from 90 kVp to 120 kVp. The in-plane pixel spacing was between 0.60 and 0.96 for both x and y direction, and the slice thickness was 1.0mm ~ 2.0mm. **More details were described in supplementary material.**

Nodule Classification

Given a CT image, segmentations of lungs and nodules were first obtained automatically using U-net [24] based segmentation models. Experienced annotators then reviewed the results and manually correct the masks if needed. Next, the segmentation and CT images were used to train the nodule classification. In this work, we compared radiomics-based approach and two deep learning-based models. The flow chart for the whole process was shown in Figure 2 A. For the radiomic model, the 3D lung nodule ROI was used as inputs. 91 radiomics features extracted from nodule mask and image patches [25] and the top important features for the classification were then selected for the final classification. The pipeline was shown in Figure 2 B. For the DL methods, we proposed one local-global model and one local model. In the local model, the 3D nodule ROI was used as input to a 3D CNN network, which used ResNet101 as the backbone [26], to learn the classification between lung cancer and cryptococcosis. However, in the local-global model two parallel CNN paths were adopted to learn both the local features around nodule ROI and the global features from the whole lung. The lung mask was used to remove all images but lungs. In order to achieve high performance, our deep learning framework used a new loss function to directly optimize AUC, which was previously reported to be effective especially for disease classification [27], **the way of adapting AUC loss to optimize the proposed DL models were detailed in the**

supplementary material. Transfer learning [28][29] was applied to overcome the overfitting issue of the DL model. The pipelines for both DL models were shown in Figure 2C.

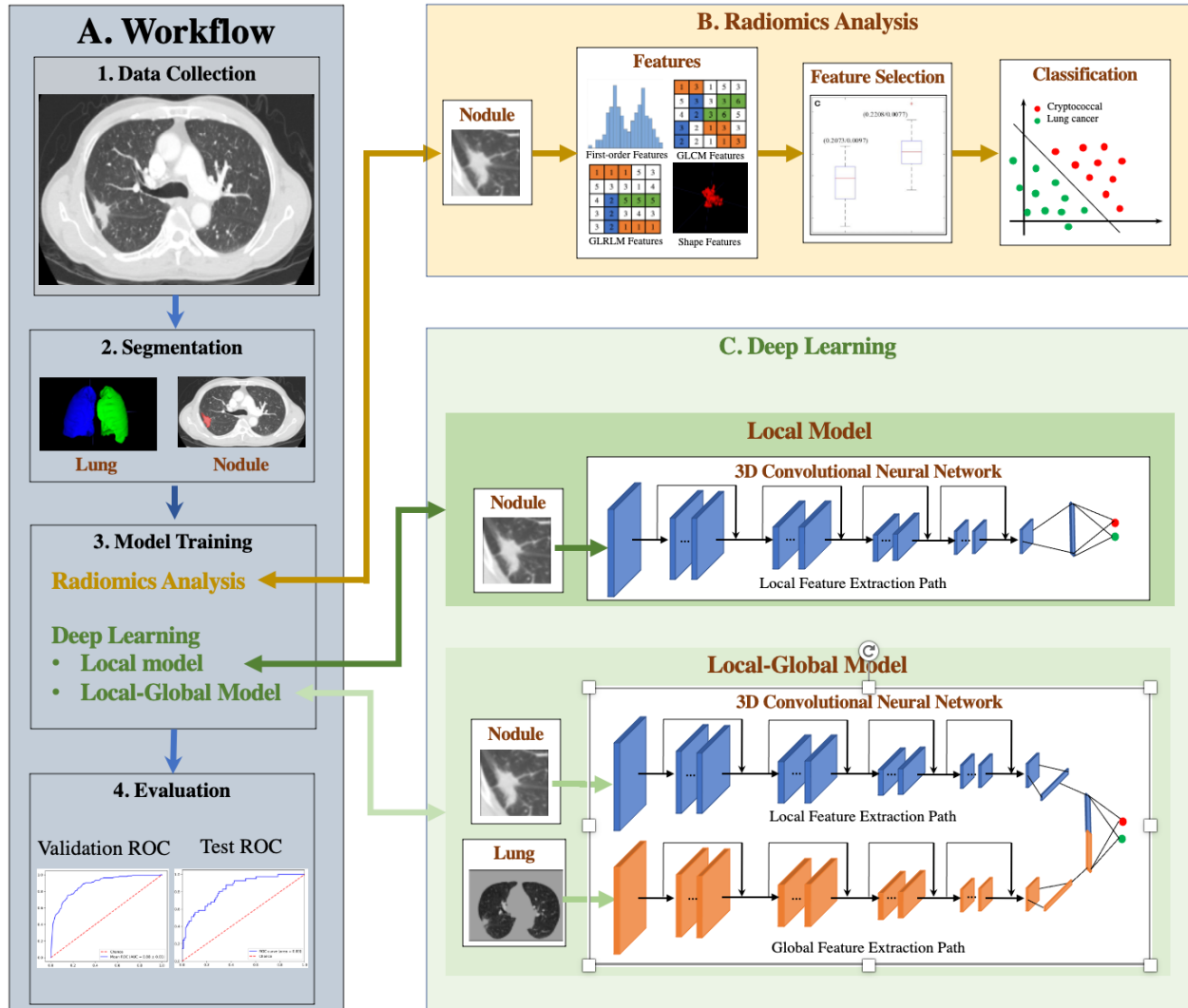


Figure 2. The method flow chart, A. Flow Chart, B. Deep Learning pipelines, C. Radiomics Pipeline.

Statistical Analysis

Baseline characteristics were presented as the count with (n) the percentage (%) for categorical variables and mean \pm standard deviation for quantitative variables. The performance of our three pipelines were firstly evaluated by five-fold cross validation on the training dataset, and the mean receiver operating characteristic (ROC) curve was plotted and the area under the curve (AUC) was calculated. The 95% confidence interval (CI) was calculated. Then the model with the highest

AUC was selected to evaluate on the test dataset to generate the ROC curve and AUC. Statistical analysis was performed using Python (version 3.6.9). A two-sided p value less than 0.05 was considered to be statistically significant. The optimal sensitivity and specificity for the cryptococcal granulomas were selected from the ROC curve and reported.

Results

Baseline Patient and nodule characteristics

Table 2 listed the baseline patient information and nodule characteristics for all patients and the two subgroups. The cohort contained 169 male patients and 126 female patients with the mean age of 56.2 years with a standard deviation of 10.9 years. Patients with lung cancers were slightly older than ones with cryptococcosis (57.1 ± 10.5 years vs 54.0 ± 11.7 years) but no significant difference was observed ($p=0.33$). The lung volumes and nodule sizes were also calculated. The average lung volume was 3534.8 milliliter (ml) with a standard deviation of 1138.2 ml. The average nodule diameter was 25.5 mm with a standard deviation of 8.0 mm. The lung volume was slightly larger for patients with lung cancers than that for patients with cryptococcosis (3552.8 ± 1184.6 ml vs 3491.9 ± 1017.8 ml). The average value of the equivalent nodule diameter (equivalent diameter assuming that a nodule is a sphere) was 25.5 mm with a standard deviation of 8.0 mm. Cryptococcosis nodules were slightly larger than malignant lung cancer nodules (equivalent diameter: 26.8 ± 11.0 mm vs 24.6 ± 6.4). However, no significant differences were observed for lung volume ($p=0.66$), nodule volume ($p=0.14$) and equivalent nodule diameter ($p=0.23$) between the two diseases types.

Table 1. Patient demographic information and nodule characteristics for all patients and the cryptococcosis and lung cancer groups.

	All Patients (n=295)	Cryptococcosis (n=87)	Lung Cancer (n=208)	<i>p</i>
Age (year)	56.2 ± 10.9	54.0 ± 11.7	57.1 ± 10.5	0.33
Male	169 (57.2%)	52 (59.8%)	117 (56.3%)	N/A
Lung volume (ml)	3534.8 ± 1138.2	3491.9 ± 1017.8	3552.8 ± 1184.6	0.66
Nodule volume (ml)	11.8 ± 16.6	16.4 ± 26.0	10.1 ± 10.9	0.14
Equivalent nodule diameter (mm)	25.5 ± 8.0	26.8 ± 11.0	24.6 ± 6.4	0.23

Classification Performance

The performance of classification between cryptococcosis and malignant lung cancer nodules was evaluated using the five-fold cross validation in the training dataset firstly, then the model with the best AUC was selected and evaluated on the test dataset. The results were plotted in Figure 3.

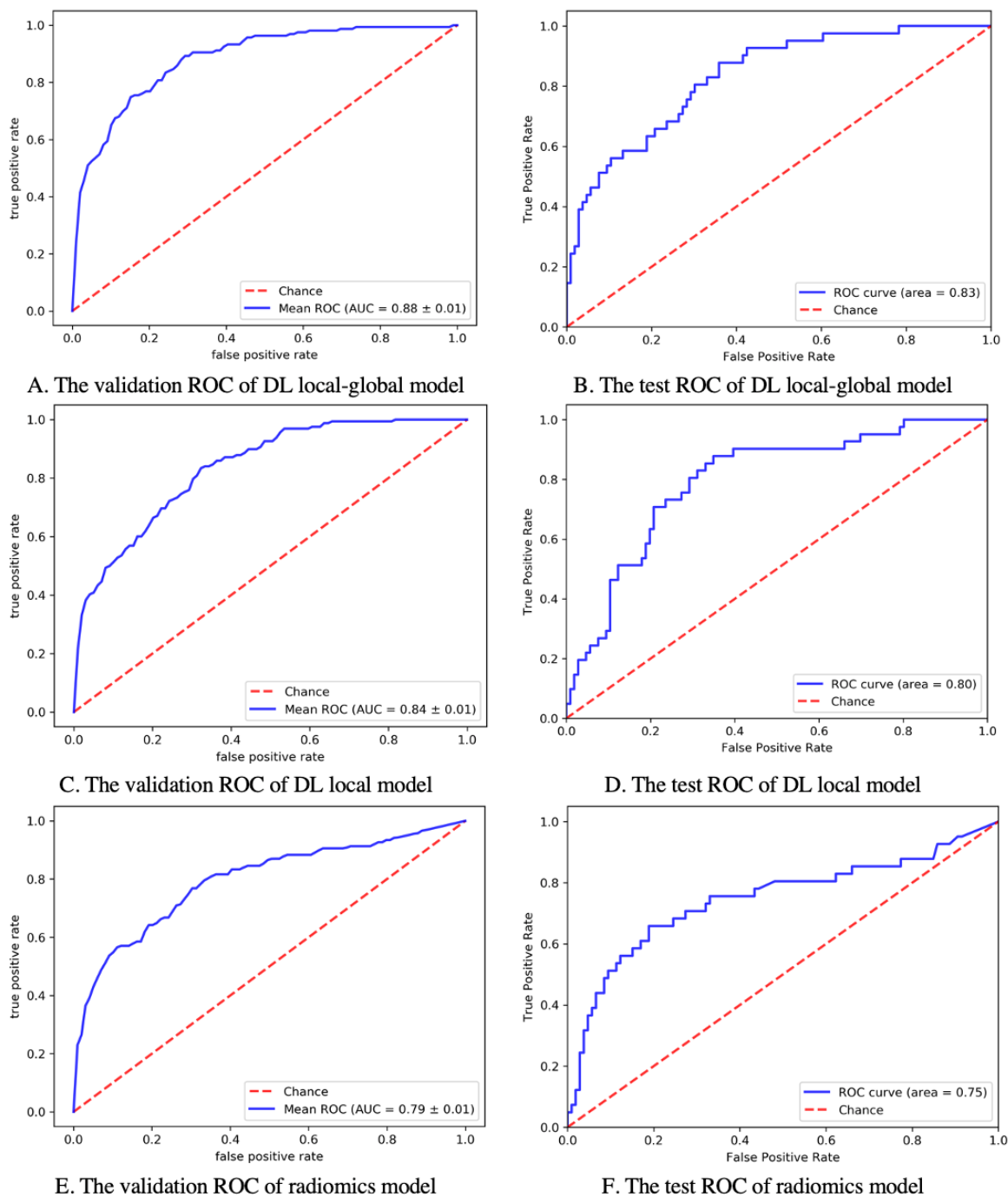


Figure 3. The validation and test ROC curves for each of three classification pipelines. A. The validation ROC of DL local-global model. B. The test ROC of DL local-global model. C. The

evaluation ROC of DL local model. D. The test ROC of DL local model. E. The validation ROC of radiomics model. F. The test ROC of radiomics model.

It can be seen that the DL local-global model yielded the highest validation AUC value (AUC=0.88) as well as the highest test AUC value (AUC=0.83) among all the three models, demonstrating that the global information extracted from lungs could be useful in distinguishing cryptococcosis nodules from lung cancer ones. Additionally, the deep learning local model had higher validation AUC value as well as the test AUC value than the radiomic model (validation AUC: local model 0.84 vs radiomic model 0.79, test AUC: local model 0.80 vs radiomic model 0.75).

Table 4 ,5 further listed the optimal sensitivity and specificity for the three models on the training dataset and the test dataset. Again, the deep learning local-global model achieved the best validation sensitivity (0.85) and specificity (0.85), and the best test sensitivity (0.81) and specificity (0.80) compared to the other two local models. The DL local model yielded higher performance than the radiomics on both training dataset and the test dataset.

Table 2. The Classification Performance of the Tree Pipelines

	threshold	Sensitivity	Specificity	Accuracy
dual-path DL	0.39	0.85	0.85	0.85
single-path DL	0.525	0.82	0.80	0.80
radiomics	0.35	0.79	0.75	0.76

Table 3. The Classification Performance of the Three Pipelines

	threshold	Sensitivity	Specificity	Accuracy
dual-path DL	0.38	0.81	0.80	
single-path DL	0.58	0.78	0.75	
radiomics	0.27	0.75	0.71	

Discussion

In this retrospective study, we investigated the possibility to differentiate solitary benign cryptococcal nodules from malignant lung cancer nodules based on chest CT. Both radiomics and deep learning methods were proposed, and the performance was evaluated. For the deep learning, we proposed one local model and one local-global one to evaluate the benefits by adding the global information. The results showed that the deep learning local-global model yielded the highest validation AUC value (AUC=0.88) as well as the highest test AUC value (AUC=0.83) among all

the three models, demonstrating that the global information extracted from lungs could be useful in distinguishing cryptococcosis nodules from lung cancer ones. Additionally, the deep learning local model had slightly higher AUC value than the radiomic features for both validation and test AUC (validation AUC : local model 0.84 vs radiomic 0.79, test AUC: local model 0.80 vs. radiomics 0.75).

CT imaging is an effective non-invasive way in evaluating various lung diseases and is especially widely used for lung cancer screening for early diagnosis of malignant lung cancer. Cryptococcosis, which is an infection caused by yeast like fungus, might present a single large nodule on CT with similar imaging manifestations, leading to a high misdiagnose rate into lung cancer. Figure 4 showed several examples of cryptococcosis and malignant lung cancer nodules, from which we could see that it is very challenging to correctly differentiate these two types of nodules based on the features related the nodule ROI itself including shape, sizes and etc. However, since cryptococcosis is one type of lung infections causing pneumonia, some additional pneumonia CT findings might be present in lung regions. Thus, we hypothesized that it might be able to improve the performance by incorporating the global information into the local prediction model.

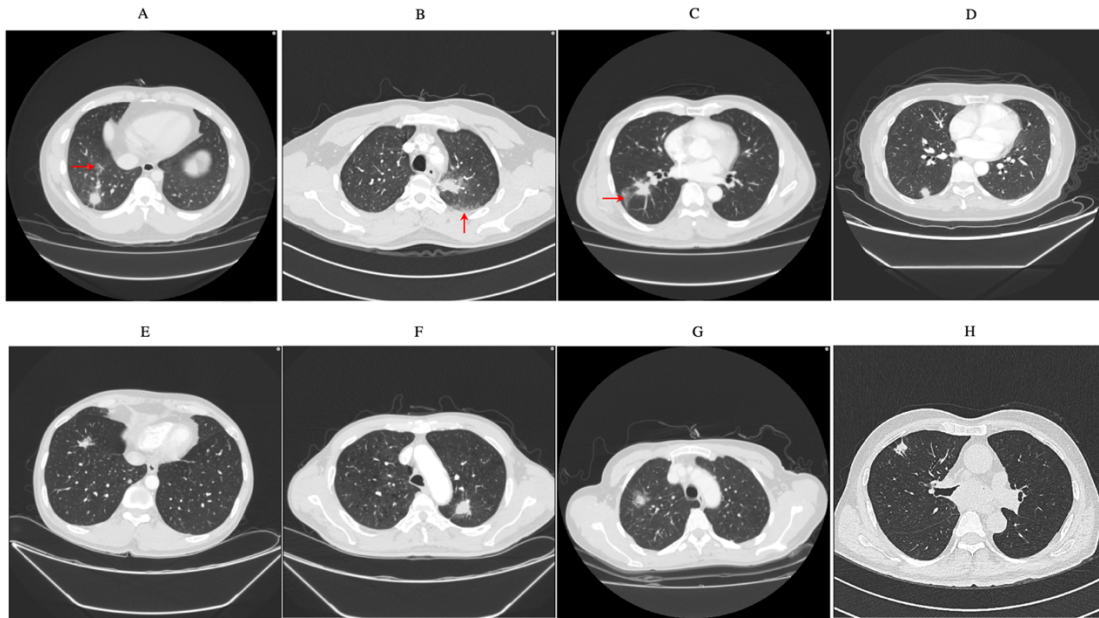


Figure 4. Several examples from the cohort. A-D: cryptococcosis nodule; E-H: malignant lung cancer nodules. The red arrows in A-C indicated possible other abnormal lung CT findings.

Various studies have been used in evaluating lung nodule especially with the majority of approaches using intramodular features. For example, Alilou et al [23] used shape-based geometric features to distinguish adenocarcinoma from granulomas and yielded an AUC of 0.72. In order to account for heterogeneity patterns in the immediate vicinity around nodules, Beig et al [24] extracted perinodular radiomic features in addition to intranodular ones to classify adenocarcinomas from granuloma and showed the improved performance compared against intranodular features alone. The authors also presented the performance of a 2D CNN based model, which used the nodule ROI as input, and showed compared results with the intramodular radiomic model. However, the perinodular region was defined based on the distance from the nodule boundary and the top radiomics features needed to be pre-selected to achieve satisfactory performance. In the situation like cryptococcosis pneumonia, the CT finding could be presented in other parts of lungs not adjacent to nodules. Thus, in this study we thus proposed a local and global deep learning model. The local information was extracted from nodule ROI images while the global information was from the whole lung. The results showed that this hybrid model yield the improved performance than the methods based on local nodule only.

Our study also has limitations. First this is a retrospective study with limited data samples of cryptococcosis. Cryptococcosis is not common diseases and the requirements of pathological confirmation for training and evaluation further reduced the sample size. To the best of our knowledge, our study included relatively large number of pathologically confirmed samples. Secondly, the current study only focused on the solitary nodules for both cryptococcosis and lung cancer. If there are two or more nodules present in the lung, it requires additional work to combine the information across the different nodules. Finally, this study only included data from a single site and it is desirable to comprehensively evaluate its performance in a multi-center study.

In conclusion, we evaluated the ability of radiomics and deep learning models in classifying solitary benign cryptococcal nodules and malignant lung cancer nodules based on chest CT. Deep learning model incorporating the global information could potentially improve the classification performance.

Reference:

- [1] F. Bray, J. Ferlay, I. Soerjomataram, Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, (2018) 394–424.
<https://doi.org/10.3322/caac.21492>.
- [2] C. Zappa, S.A. Mousa, Non-small cell lung cancer: current treatment and future advances, *Transl. Lung Cancer Res.* 5 (2016) 288.
- [3] Y. Zhang, N. Li, Y. Zhang, H. Li, X. Chen, S. Wang, X. Zhang, R. Zhang, J. Xu, J. Shi, others, Clinical analysis of 76 patients pathologically diagnosed with pulmonary cryptococcosis, *Eur. Respir. J.* 40 (2012) 1191–1200.
- [4] Z. Hu, J. Chen, J. Wang, Q. Xiong, Y. Zhong, Y. Yang, C. Xu, H. Wei, Radiological characteristics of pulmonary cryptococcosis in HIV-infected patients, *PLoS One.* 12 (2017) e0173858.
- [5] L.-W. Gao, A.-X. Jiao, X.-R. Wu, S.-Y. Zhao, Y. Ma, G. Liu, J. Yin, B.-P. Xu, K.-L. Shen, Clinical characteristics of disseminated cryptococcosis in previously healthy children in China, *BMC Infect. Dis.* 17 (2017) 359.
- [6] M. Guarana, J.E. Vidal, M. Nucci, Cryptococcosis in Patients with Hematologic Diseases, *Curr. Fungal Infect. Rep.* 12 (2018) 187–194.
- [7] W. Cao, C. Jian, H. Zhang, S. Xu, Comparison of clinical features and prognostic factors of cryptococcal meningitis caused by *Cryptococcus neoformans* in patients with and without pulmonary nodules, *Mycopathologia.* 184 (2019) 73–80.
- [8] J. Qu, X. Zhang, Y. Lu, X. Liu, X. Lv, Clinical analysis in immunocompetent and immunocompromised patients with pulmonary cryptococcosis in western China, *Sci. Rep.* 10 (2020) 1–11.
- [9] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, *Radiology.* 278 (2016) 563–577.
- [10] H.J.W.L. Aerts, E.R. Velazquez, R.T.H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, others, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, *Nat. Commun.* 5 (2014) 1–9.
- [11] Q. Weng, L. Zhou, H. Wang, J. Hui, M. Chen, P. Pang, L. Zheng, M. Xu, Z. Wang, J. Ji, A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part-solid nodules, *Clin. Radiol.* 74 (2019) 933–943.
- [12] X. Yang, J. He, J. Wang, W. Li, C. Liu, D. Gao, Y. Guan, CT-based radiomics signature for differentiating solitary granulomatous nodules from solid lung adenocarcinoma, *Lung Cancer.* 125 (2018) 109–114.
- [13] M. Cong, H. Feng, J.-L. Ren, Q. Xu, L. Cong, Z. Hou, Y. Wang, G. Shi, Development of a predictive radiomics model for lymph node metastases in pre-surgical CT-based stage IA non-small cell lung cancer, *Lung Cancer.* 139 (2020) 73–79.
- [14] W. Tu, G. Sun, L.I. Fan, Y. Wang, Y. Xia, Y. Guan, Q. Li, D. Zhang, S. Liu, Z. Li, Radiomics signature: a potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology, *Lung Cancer.* 132 (2019) 28–35.
- [15] C. Dennie, R. Thornhill, V. Sethi-Virmani, C.A. Souza, H. Bayanati, A. Gupta, D. Maziak, Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules, *Quant. Imaging Med. Surg.* 6 (2016) 6.
- [16] S. Hawkins, H. Wang, Y. Liu, A. Garcia, O. Stringfield, H. Krewer, Q. Li, D. Cherezov, R.A. Gatenby, Y. Balagurunathan, others, Predicting malignant nodules from screening CT scans, *J. Thorac. Oncol.* 11 (2016) 2120–2128.
- [17] S.S. Siegelman, E.A. Zerhouni, F.P. Leo, N.F. Khouri, F.P. Stitik, CT of the solitary pulmonary nodule, *Am. J. Roentgenol.* 135 (1980) 1–13.
- [18] G. Kang, K. Liu, B. Hou, N. Zhang, 3D multi-view convolutional neural networks for lung nodule classification, *PLoS One.* 12 (2017) e0188290.
- [19] P. Sahu, D. Yu, M. Dasari, F. Hou, H. Qin, A lightweight multi-section CNN for lung nodule classification and malignancy estimation, *IEEE J. Biomed. Heal. Informatics.* 23 (2018) 960–968.
- [20] J. Gong, J. Liu, W. Hao, S. Nie, B. Zheng, S. Wang, W. Peng, A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images, *Eur. Radiol.* (2019) 1–9.
- [21] D. Ardila, A.P. Kiraly, S. Bharadwaj, B. Choi, J.J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, others, End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, *Nat. Med.* 25 (2019) 954–961.
- [22] D.R. Baldwin, J. Gustafson, L. Pickup, C. Arteta, P. Novotny, J. Declerck, T. Kadir, C. Figueiras, A. Sterba,

- A. Exell, others, External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules, *Thorax*. 75 (2020) 306–312.
- [23] R. Dey, Z. Lu, Y. Hong, Diagnostic classification of lung nodules using 3D neural networks, in: 2018 IEEE 15th Int. Symp. Biomed. Imaging (ISBI 2018), 2018: pp. 774–778.
 - [24] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Int. Conf. Med. Image Comput. Comput. Interv., 2015: pp. 234–241.
 - [25] J.J.M. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R.G.H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, H.J.W.L. Aerts, Computational radiomics system to decode the radiographic phenotype, *Cancer Res.* 77 (2017) e104--e107.
 - [26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016: pp. 770–778.
 - [27] L. Yan, R.H. Dodier, M. Mozer, R.H. Wolniewicz, Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic, in: Proc. 20th Int. Conf. Mach. Learn., 2003: pp. 848–855.
 - [28] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding transfer learning for medical imaging, in: Adv. Neural Inf. Process. Syst., 2019: pp. 3347–3357.
 - [29] S. Chen, K. Ma, Y. Zheng, Med3d: Transfer learning for 3d medical image analysis, *ArXiv Prepr. ArXiv1904.00625*. (2019).

Supplementary Material

The Details of the Scanner Parameters

714 contrast CT images with slice thickness less than or equal to 2.0 mm were acquired from scanners of five different manufactures (PHILIPS, TOSHIBA, SIEMENS, GE, UIH) under six kVps (80, 90, 100, 120, 140, 150). Among them, more than 85% scans (672/714) were taken under 120 kVp, the slice thickness of more than 55% scans (395/714) was 2.0 mm. The distribution of different kVps and slice thicknesses across five manufactures were detailed in the Table S1 below.

Table S4. The details of the scanner parameters

Vendor	kVp	Slice Thickness	Total No.
PHILIPS	120 (212)	134 (1.0 mm), 78 (2.0 mm)	212
TOSHIBA	120 (286)	4 (1.0 mm), 282 (2.0 mm)	286
SIEMENS	100(8), 80 (9), 150 (1), 120(2), 90 (16)	30 (1.25 mm) 6 (2.0 mm)	36
GE	120 (168), 140 (8)	149 (1.25 mm), 27(2.0 mm)	176
UIH	120 (4)	2 (1.0 mm), 2 (2.0 mm)	4

AUC loss Function for Optimizing 3D ResNet

Recently, the deep convolutional neural network (CNN) has demonstrated its advantage in the image feature extraction. 3D ResNet is a famous deep CNN with the residual block, and widely applied in many classification tasks in medical image analysis[1,2]. With the residual block, the 3D ResNet can go very deep without increasing the model parameter size dramatically so that it can fit into the graphics processing unit memory to obtain the robust, distinguishable high-level imaging features. Such features can improve the classification performance. For the classification task, the ROC curve, and the area under the ROC curve (AUC) are appropriate performance metrics. The 3D ResNet are often optimized by minimizing the cross-entropy loss in the classification task. However, it has been demonstrated that minimizing cross-entropy does not

necessarily maximize AUC[3]. To explicitly optimize the AUC through the loss function, the AUC loss was developed as follows:

$$AUC\ Loss = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} R(pos_i, neg_j) \quad (1)$$

$$R(pos_i, neg_j) = \begin{cases} (pos_i - neg_j - \gamma)^2 & : pos_i - neg_j \leq \gamma \\ 0 & : Otherwise \end{cases} \quad (2)$$

In equation (1), M denotes the number of the positive samples in the dataset, and N denotes the number of negative samples in the dataset. pos_i ($0 \leq pos_i \leq 1$) indicates the output of the 3D ResNet classification model for i -th positive sample while neg_j ($0 \leq neg_j \leq 1$) indicates the output of the model for j -th negative sample. For each pair of positive and negative outputs, the single pair AUC loss value is calculated according to equation (2). More specifically, the loss will be calculated when pos_i is smaller than neg_j or larger than neg_j when a predefined threshold γ ($0 \leq \gamma \leq 1$). The AUC loss is obtained by accumulating each pair loss together. In the study, γ was set to 0.2 based on experiments results. Finally, the DL based models' parameters were optimized by minimizing AUC loss through gradient descent method. Intuitively, any positive output would be larger than any negative output with a large enough margin if the classifier is perfect. If the output of a negative sample is larger than the output of a positive sample, or the output of a positive sample is not large enough, the model parameters will be tuned by the AUC loss of such a pair.

- [1] X. Yin, Q. Zhao, J. Liu, W. Yang, J. Yang, G. Quan, Y. Chen, H. Shu, L. Luo, J.-L. Coatrieux, Domain progressive 3D residual convolution network to improve low-dose CT imaging, *IEEE Trans. Med. Imaging*. 38 (2019) 2903–2913.
- [2] E. Shim, J.Y. Kim, J.P. Yoon, S.-Y. Ki, T. Lho, Y. Kim, S.W. Chung, Automated rotator cuff tear classification using 3D convolutional neural network, *Sci. Rep.* 10 (2020) 1–9.
- [3] W. Gao, Z.-H. Zhou, On the Consistency of AUC Pairwise Optimization., in: *IJCAI*, 2015: pp. 939–945.