The Value of CT Radiomics Analysis of Cavity Characteristics In Differentiating Nontuberculous Mycobacterium Pulmonary Disease From Pulmonary Tuberculosis

Qinghu Yan

Shandong Province Chest Hospital

Jia Cui

Shandong Province Chest Hospital

Wenlong Zhao

Shandong Province Chest Hospital

Liping Zuo

Qilu Hospital of Shandong University

Dongdong Wang

Qilu Hospital of Shandong University

Wuzhang Wang

Shandong Province Chest Hospital

Xiangfei Cai

Huiying Medical Technology (Beijing) Co., Ltd

Dexin Yu (✉ yudexin0330@sina.com)

Qilu Hospital of Shandong University

Research Article

Keywords: Radiomics, Cavity, CT, Nontuberculous Mycobacteria, Pulmonary tuberculosis

Posted Date: May 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-468820/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The CT features and clinical manifestations of the NTM pulmonary disease are similar to those of pulmonary tuberculosis (PTB), so it is very difficult to identify these two diseases simply via CT or clinical features.

Methods: From February 2013 to March 2018, 73 patients of NTM pulmonary diseases and 69 patients of PTB with the cavity in Shandong Province Chest Hospital and Qilu Hospital of Shandong University were retrospectively analyzed. A double-blind assessment and manual delineation of 300 cavities as the region of interesting (ROI) from chest CT images were performed by 2 experienced radiologists. 80% of cavities were allocated to the training set and 20% to the validation set using a random number generated by a computer. 1409 radiomics features extracted from the Huiying Radcloud platform were used to analyze the two kinds of diseases' CT cavity characteristics. Feature selection was performed using analysis of variance (ANOVA) and least absolute shrinkage and selection operator (LASSO) methods, and six supervised learning classifiers (KNN, SVM, XGBoost, RF, LR, and DT models) were used to analyze the features. The Receiver Operating Characteristic (ROC), sensitivity and specificity were used to evaluate the model's performance.

Results: 29 optimal features were selected by the variance threshold method, K best method, and Lasso algorithm. and the ROC curve values are obtained. In the training set, the AUC values of the six models were all greater than 0.97, 95% CI were 0.95-1.00, the sensitivity was greater than 0.92, and the specificity was greater than 0.92. In the validation set, the AUC values of the six models were all greater than 0.84, 95% CI were 0.76-1.00, the sensitivity was greater than 0.79, and the specificity was greater than 0.79.

Conclusion: The radiomics features extracted from cavity on CT images can provide effective proof in distinguishing the NTM pulmonary disease from PTB, and the radiomics analysis shows a more accurate diagnosis than the radiologists.

Background

The incidence of nontuberculous mycobacteria (NTM) has been increasing rapidly worldwide in recent years (1-2). NTM, usually involving the lung, has the characteristics such as slow growth, long course, and high resistance to first-line anti-tuberculosis drugs, which caused a low total care rate.(3). In order to improve the treatment effect, the accuracy of the diagnosis before treatment is crucial. However, the CT features and clinical manifestations of the NTM pulmonary disease are similar to those of pulmonary tuberculosis (PTB), so it is very difficult to identify these two diseases simply via CT or clinical features. At present, the main methods to distinguish them include the culture of mycobacterium in sputum and the identification of the species. However, these methods are time-and resource-consuming and require higher laboratory standards, which can not be provided in many local hospitals (4). Moreover, even culture result is positive, further identify the species is also needed. Some studies (5-6) have shown that there usually were some imaging changes of NTM pulmonary disease, particularly the cavity of the lesion

on chest CT images, and the cavity is also a common imaging feature of chronic PTB (7). Unfortunately, most cavities on CT of this two kinds of diseases are difficult to be an effective identification points for the radiologists. In recent years, radiomics, mining data features through high-throughput extraction, shows a great potential in the diagnosis and differentiation of lung diseases, such as pulmonary nodules and early lung cancer (8,9), which may provide a possible method to identify the NTM pulmonary disease from PTB. In addition, the pathological difference in the cavity between two kinds of diseases have been proved (10), which increases the identification possibility for them. In this piece, it is the first study to investigate the value of CT radiomics analyzing cavity in identification of the NTM pulmonary disease and PTB. Objective to explore the value of this technique in the differential diagnosis of two diseases, and to provide a new and simple diagnostic method for clinical treatment.

Materials And Methods

2.1 Patient information

The study retrospectively analyzed 89 patients with NTM pulmonary disease and other 100 cases of patients with PTB with cavity (11) in CT imaging in the Shandong Provincial Chest Hospital and Qilu Hospital of Shandong University from February 2013 to March 2018. The content of this study was approved by the ethics committee of Shandong University. All patients with lung changes underwent a plain chest scan using Philips 64-slice spiral CT before treatment, and the quality correction was qualified before the CT scan. The patient with NTM pulmonary disease or PTB lay down in supine position, and CT scan from the tip of the lungs to the level of the thoracic cavity was carried out when holding his breath. CT scanning parameters were as following: the diameter of the inspected detector, 64×0.625 mm; the rotation time, 0.5 s; the pitch, 1.375; the tube voltage, 120 kV; the tube current, 250 ~ 400 mA, which was modulated using an automatic tube current; and the thickness of the slice, 5 mm.

The inclusion criteria: (1) All cases of patients with NTM pulmonary disease were diagnosed with sputum cultured twice and the strains were identified as the same pathogenic bacteria (7,12). All cases of patients with PTB were identified by the new version of the tuberculosis diagnosis and tuberculosis classification standard issued by China National Health Service Commission on November 9, 2017. 1, Twice sputum samples of the same patient were positive for acid-fast bacilli after smearing, and were identified as mycobacterium tuberculosis complex by colloidal gold method. (2) The CT images (lung window) indicated the existence of cavity more than 3mm. (3) Corresponding anti-NTM or -PTB treatment was effective. (4) All images were met the analytical criteria of the radiomics. Exclusion criteria: (1) CT image showed the lesions without cavities. (2) Some factors might influenced the accuracy of analysis including CT images with motion artifacts, poor image quality, massive differences in scanning parameters and inconsistent slice thicknesses. (3) The bullae and other cavities without wall. (4) Corresponding treatments had no effect. (5) The clinical data showed that the patient had the possibility of mixed infection. (6,) The outer wall is difficult to be delineated because it was covered by the surrounding opacities or consolidation. After screening, 16 patients with NTM pulmonary diseases and 31 patients with PTB were excluded, A total of 73 patients with NTM pulmonary disease were enrolled in this

study, including 47 males and 26 females, with an average age of 57.3 ± 14.0 years (ranged from 23- 86 years). And other 69 patients were with PTB, including 48 males and 21 females; mean age 45.7 ± 17.9 years (ranged from 17-90 years).

2.2 Characterization of lesions

All imaging and clinical data were uploaded to a Radiance's Radcloud platform (Huiying Medical Technology Co., Ltd) for further study. Two radiologists (reader 1: Q.Y, Six years of experience, and reader 2: S.Q, Ten years of experience) with expertise in chest disease were invited to interpret and delineate the cavities on lung-window CT images (W 1500, L -500). The two readers were both blinded to the clinical and pathogenic information. The first and last slices were excluded in order to eliminate errors caused by the volume effect, then the reader 1 manually sketched the ROI by delineating the outline of the entire cavity on all contiguous slices. The contour line was drawn on the outer edge of the cavity and should completely include the cavity and its wall. If there were ground-glass opacities or consolidation partly next to the cavity, the outline was drawn along the existing outer wall on referring to its thickenss and the shape of inner wall. If there were multiple cavities in the patient's lungs, each cavity was individually outlined as an independent ROI; and if two or more lung cavities were connected or multilocular, they would be treated as one cavity to be drawn the outer wall. The small cavity with a diameter less than 0.5cm was not drawn because of possible delineating error. A total of 301 ROIs were manually mapped out from CT images of 142 patients. Afterwards, reader 2 with senior professional title then reviewed all the delineation. If the difference between the two readers was greater than or equal to 5%, the senior reader determined the cavity ROI boundary (13). The drawn contour lines of the cavity on the lung-window CT images were shown in Fig. 1 (NTM pulmonary disease) and Fig. 2 (PTB).

2.3 Features extraction

A total of 1409 quantitative imaging features were extracted from ROIs with the RadCloud platform. These features were divided into three groups: Group 1 (intensity statistics features) consisted of 270 features that quantitatively delineated the distribution of voxel intensities within the CT images through commonly used and basic metrics. Group 2 (shape- and size-based features) contained 16 three-dimensional features that reflected the shape and size of the region. Group 3 (texture features) calculated by grey level run-length matrices (GLRLM), grey level co-occurrence texture matrices (GLCM), gray level size zone matrix (GLSZM), neighbouring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM), and 1123 textural features that could quantify region heterogeneity differences were classified into group 3 (texture features).

2.4 Features selection

As described above, high-throughput image features may be computed. However, not all the features were useful for a particular task. Therefore, in order to obtain the most relevant features, feature selection for the specific task was a necessary steps. Analysis of variance (ANOVA) and least absolute shrinkage and selection operator (LASSO) were used to remove redundant feature. ANOVA belongs univariate feature

selection method, using P value ≤ 0.05 as the significant differences in features between the two lesions. For the LASSO method, L1 regularizer was used as the cost function, the error value of cross validation was 5, and the maximum number of iterations was 1000.

2.5 Statistical analysis

The statistical analysis was performed in the Radcloud platform. After feature selection, a total of 66 features identified were significantly correlated to differentiating NTM pulmonary disease from PTB. Based on the selected features, there were several supervised learning classifiers available for classification analysis. In this study, the radiomics-based models were constructed with 6 classifiers, k-NearestNeighbor (KNN), Support Vector Machin (SVM), eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Logistic Regression (LR) and Decision tree(DT), and the validation set was used to evaluate the effectiveness of these models. To assess the predictive performance, the receiver operating characteristic (ROC) curve and the area under curve (AUC) were used both in training set and validation set respectively. And four indicators including P (precision = true positives / (true positives+ false positives)), R (recall = true positives / (true positives+ false negatives)), f1-score (f1-score = $P \times R \times 2 / (P + R)$), support (total number in test set) to evaluate the performance of classifier in this study.

Results

Firstly, 476 features were selected from 1409 features using the variance threshold method (Fig. 3). Then 349 features were selected with the ANOVA method (Fig.4). Finally, we selected 29 optimal features with LASSO algorithm (Fig. 5).When training with KNN SVM, XGBoost, RF, LR and DT classifier, the AUC of training set and validation set were shown in Fig. 6- 11.

Tab. 1 ROC results with KNN, SVM, XGBoost, RF, LR and DT classifiers of training set

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	NTM	0.9886909810573932	0.96 - 1.00	0.92	0.96
	TB	0.9886909810573933	0.96 - 1.00	0.96	0.92
SVM	NTM	0.979926491376873	0.95 - 1.00	0.95	0.96
	TB	0.9799264913768732	0.95 - 1.00	0.96	0.95
XGBoost	NTM	0.9998586372632173	0.99 - 1.00	0.98	0.99
	TB	0.9998586372632174	0.99 - 1.00	0.99	0.98
RF	NTM	0.9999646593158045	0.99 - 1.00	0.99	1
	TB	0.9999646593158044	0.99 - 1.00	1.00	0.99
LR	NTM	0.9868532654792197	0.96 - 1.00	0.95	0.95
	TB	0.9868532654792197	0.96 - 1.00	0.95	0.95
DT	NTM	1	1.00 - 1.00	1.00	1
	TB	1	1.00 - 1.00	1.00	1

Tab. 2 ROC results with six classifiers of validation set

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	NTM	0.987012987012987	0.93 - 1.00	0.94	0.96
	TB	0.987012987012987	0.93 - 1.00	0.96	0.94
SVM	NTM	0.9837662337662338	0.91 - 1.00	0.97	0.82
	TB	0.9837662337662338	0.91 - 1.00	0.82	0.97
XGBoost	NTM	0.9848484848484849	0.94 - 1.00	0.97	0.96
	TB	0.9848484848484849	0.94 - 1.00	0.96	0.97
RF	NTM	0.9886363636363635	0.93 - 1.00	0.94	0.96
	TB	0.9886363636363636	0.93 - 1.00	0.96	0.94
LR	NTM	0.9751082251082253	0.91 - 1.00	0.94	0.93
	TB	0.9751082251082251	0.91 - 1.00	0.93	0.94
DT	NTM	0.8474025974025974	0.76 - 0.94	0.91	0.79
	TB	0.8474025974025973	0.76 - 0.94	0.79	0.91

Six classifiers were used to analyze the characteristics of radiomic AUC, 95% CI, sensitivity and specificity of training set and verification set were shown in Tab.1 and Tab. 2. In the training set, the AUC values of the six models were all greater than 0.97, 95% CI were 0.95-1.00, the sensitivity was greater than 0.92, and the specificity was greater than 0.92. In the validation set, the AUC values of the six models were all greater than 0.84, 95% CI were 0.76-1.00, the sensitivity was greater than 0.79, and the specificity was greater than 0.79. It was found from the table data that no matter the training set or the truth verification set, the AUC value was very high, which was greater than 0.84. The AUC of DT classifier was the lowest, 95% CI was 0.76-0.94, sensitivity and specificity was greater than 0.79.

Tab. 3 The results of four indicators -Precision, Recall, F1-score, Support in training set

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
TB	Precision	0.90	0.95	0.98	0.99	0.94	1.00
	Recall	0.96	0.96	0.99	1.00	0.95	1.00
	F1-score	0.93	0.95	0.99	1.00	0.95	1.00
	Support	108	108	108	108	108	108
NTM	Precision	0.97	0.97	0.99	1.00	0.96	1.00
	Recall	0.92	0.95	0.98	0.99	0.95	1.00
	F1-score	0.94	0.96	0.99	1.00	0.96	1.00
	Support	131	131	131	131	131	131

Tab. 4 The results of four indicators -Precision, Recall, F1-score, Support in validation set

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
TB	Precision	0.93	0.96	0.96	0.93	0.93	0.88
	Recall	0.96	0.82	0.96	0.96	0.93	0.79
	F1-score	0.95	0.88	0.96	0.95	0.93	0.83
	Support	28	28	28	28	28	28
NTM	Precision	0.97	0.86	0.97	0.97	0.94	0.83
	Recall	0.94	0.97	0.97	0.94	0.94	0.91
	F1-score	0.95	0.91	0.97	0.95	0.94	0.87
	Support	33	33	33	33	33	33

In the training set, the accuracy of the six models was greater than 0.90, the recall rate was greater than 0.92, the F1 score was greater than 0.93, and the support were 108 and 131(Tab. 3). In the validation set,

the accuracy of the six models was greater than 0.83, the recall rate was greater than 0.79, the F1 score was greater than 0.83, and the support were 28 and 33(Tab. 4).

Discussion

NTM pulmonary disease refers to lung diseases caused by mycobacteria except mycobacterium tuberculosis complex and mycobacterium leprae. These bacteria exist widely in water and soil and a total of 156 types of NTM have been found in former study (3,14-15), of which only a small number of bacteria can cause disease. The incidence and prevalence of NTM pulmonary disease are increasing in some countries and regions (16-19). In some areas, they are even higher than those of PTB (3). The identification of NTM pulmonary disease and PTB requires pathogenic testing, but it is a complex step. And the simple and non-invasive chest CT imaging has a very important role in diagnosis and differential diagnosis. CT of NTM pulmonary disease shows a variety of signs such as cavity, bronchiectasis, and bronchiectasis, which makes it difficult to distinguish from PTB (20,21). Among them, cavity is a more common imaging manifestation in NTM pulmonary disease (22,23). There are plenty literature results on the comparison of cavity types of pulmonary disease, and the opinions are widely divergent (24,25). Some studies (26) had shown that the cavities of NTM pulmonary disease are more located near the pleura of the lungs than those of PTB. Other studies (22) pointed out that NTM pulmonary disease has more thin-walled cavities than PTB, while some studies suggest that there is no statistical difference in cavity wall results for the two diseases(5,23). These shown that there was still controversy in the use of conventional CT to identify the cavity features, especially its practical value.

In recent years, radiomics has developed rapidly, which converts medical images into high-dimensional images, and obtains quantifiable data features through quantitative high-throughput analysis, and then performs data analysis to provide clinical support. Radiomics has shown a very important role in the diagnosis and differential diagnosis of clinical pulmonary diseases, such as screening of small pulmonary nodules, diagnosis of lung cancer, guidance of the treatment of pulmonary disease, and prognosis prediction (26-28). It also shows good application in the diagnosis of some common infectious diseases (29,30). However, there are relatively few studies on diffuse pulmonary disease (31,32), and we assume the main reason is difficult to select and delineate the target area. The imaging of NTM pulmonary disease is complex and variable, but the cavity in the lungs is similar with PTB. In addition, studies have shown that cavity is associated with mortality (33,34), disease progression, and clinical treatment failure (35,36), so international guidelines for the diagnosis and treatment of NTM pulmonary disease indicate that patients with cavity need active treatment (37). Therefore, in this study, cavities were used as target areas for radiomics analysis, and radiomics features obtained from cavities were used to explore the value of identifying the disease from PTB.

In this study, the cavities of NTM pulmonary disease and PTB were used as potential areas of interest for delineation. A total of 29 radiomics features were obtained from ROIs, of which 14 were texture analysis and 15 were first-order statistical feature groups. The study used six supervised learning classifiers (KNN, SVM, XGBoost, RF, and LRandomDT) to analyze the extracted lung cavity features. In the training set, the

AUC values of the six models were all greater than 0.97, 95% CI were 0.95-1.00, the sensitivity was greater than 0.92, and the specificity was greater than 0.92. In the validation set, the AUC values of the six models were all greater than 0.84, 95% CI were 0.76-1.00, the sensitivity was greater than 0.79, and the specificity was greater than 0.79. The obtained AUC values of ROC curve were all high, and their sensitivity and specificity were greater than 0.79. The study analyzed the characteristics of the three classifiers through four clinical indicators (accuracy, recall, f1 points, and support). In the training set, the accuracy of the six models was greater than 0.90, the recall rate was greater than 0.92, the F1 score was greater than 0.93, and the support were 108 and 131. In the validation set, the accuracy of the six models was greater than 0.83, the recall rate was greater than 0.79, the F1 score was greater than 0.83, and the support were 28 and 33. Therefore, the results showed that the radiomics features derived from cavities could help differentiate NTM pulmonary disease and PTB. Although some studies have found that the CT imaging characteristics of NTM pulmonary disease cavities observed through traditional clinical methods are different from those of PTB(12,22,38), the results obtained by researchers' naked eye recognition may appear subjective or empirical differences. So the effect is greatly reduced in actual clinical applications. The radiomics analysis of the cavity characteristics of pulmonary disease has the advantages of objectiveness, quantification, stability, and nonexperience dependence. Therefore, the current clinical use value is greater and the clinical application prospect is broader. It can be seen that the use of the radiomics characteristics in differentiating NTM pulmonary disease and PTB is promising, and the results may help for clinical radiologist lacking experience or difficulty in visual recognition. Therefore, this technology is a very important, non-invasive, simple and easy method to identify these two diseases.

In addition, although the pathological features of NTM pulmonary disease and tuberculosis are similar, there are some differences as following:(1) the lesions of NTM pulmonary disease have less caseous necrosis than PTB; (2) epithelial-like cells aggregation tends to be proliferative granuloma; and (3) the cavities are more common and the characteristics may be with multiple forms, multiple rooms and uneven inner walls(10). Thus, the subtle differences of these pathological features on conventional CT images are difficult to be distinguished resulting in a low level of differential diagnosis from PTB. The high-throughput extracted CT data of radiomics can reflect the internal differences of the pathological characteristics of the two diseases, which is consistent with the pathological performance, and makes up for the lack of visual observation and key information in CT images.

However, this study also has some limitations: 1. In order to ensure the homogeneity of the images, 5mm layer thickness images were used in the research process. Compared with thinner slice images, some information may be lost. 2. ROIs were delineated manually and personal subjectivity is unavoidable. 3. We just focused on cavity as research target and ignored other image features, which will lead to a lack of information and some typical patients cannot be studied. 4. Data analysis models need further improvement. The above shortcomings and deficiencies will be compensated in future research.

In conclusion, this study shows that the radiomics features based on CT imaging are effective in identifying NTM pulmonary disease cavities and PTB cavities, which has great value for the early noninvasive diagnosis and differentiation of NTM pulmonary disease with cavity.

Declarations

Ethics approval and consent to participate

All authors attest that they meet the Helsinki declaration.

The authors declare that they consent to participate.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interest

The authors declare that they have no competing interest.

Funding

Organization: Shandong health and health Committee; Entry name:Shandong medical and health science and technology development plan project; Project number:2019WS535; Project Leader: Qinghu Yan

Authors' contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Acknowledgements

The authors gratefully acknowledge the financial supports by the Foundation of Shandong health and health Committee under project number 2019WS535.

Statement for informed consent

This retrospective studies included all subjects are under 18 and informed consent obtained from a parent and/or legal guardian.

References

[1] Kendall BA,Winthrop KL.Update on the epidemiology of pulmonary nontuberculous mycobacterial infections[J].Semin Respir Crit Care Med,2013,34(1):87-94.DOI: 10.3201/eid2206.151086

- [2] Simons S, van Ingen J, Hsueh PR, et al. Nontuberculous mycobacteria in respiratory tract infections, eastern Asia [J]. *Emerg Infect Dis*, 2011, 17(3):343-349. DOI: 10.3201/eid1703.100604
- [3] Chinese Medical Association Tuberculosis Branch, Editorial Board of Chinese Journal of Tuberculosis and Respiratory Diseases. Expert consensus on the diagnosis and treatment of nontuberculous mycobacterial disease. *Chinese Journal of Tuberculosis and Respiratory Diseases*, 2012, 35 (8):572-580. DOI:10.3760/cma.j.issn.1001-0939.2012.08.006
- [4] Cao Peiming, Liu Fengfeng, Li Yan, et al. Clinical analysis of 58 cases of nontuberculous mycobacteria [J]. *Chongqing Medical*, 2014, 43 (7): 854-856. DOI: 10.3969 / j.issn.1671-8348.2014.07.032
- [5] Li Fang, HE Wei, ZHOU Xin hua, et al. The similarities and differences of HRCT imaging features between nontuberculous mycobacterial lung diseases and active pulmonary tuberculosis [J]. *The Journal of the Chinese Antituberculosis Association*, 2018, 40(5) :499-505. DOI:10.3969/j.issn.1000-6621.2018.05.012
- [6] Zhang Jingbo, Chu Haiqing. Clinical and imaging features of nontuberculous mycobacterial lung diseases [J]. *Chinese Journal of Infection and Chemotherapy*, 2016, 16(1):86-91. DOI:10.16718/j.1009-7708.2016.01.020
- [7] Guo Hui. To explore the clinical value of CT in the diagnosis and clinicopathological classification of peripheral small lung cancer [J]. *China Continuing Medical Education*, 2016, 8(21):48. DOI:10.3969 / j.issn.1674-9308.2016.21.027
- [8] Ma J, Zhou Z, Ren Y, et al. Computerized Detection of Lung Nodules through Radiomics [J]. *Med Phys*, 2017, 44(8):4148-4158.
- [9] Coroller TP, Agrawal V, Huynh E, et al. Radiomic-based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC [J]. *J Thorac Oncol*, 2017, 12(3):467-476.
- [10] Wittram C, Weisbrod GL. Mycobacterium avium complex lung disease in immunocompetent patients: radiography- CT correlation [J]. *Br J Radiol*, 2002, 75(829): 340- 344.
- [11] Chinese Medical Association Tuberculosis Branch. Guidelines for the diagnosis and management of nontuberculous mycobacteria. *Chinese Journal of Tuberculosis and Respiratory Diseases*, 2000, 23 (11): 650-653. DOI: 10.3760 / j: issn: 1001-0939.2000.11.003
- [12] Daley CL, Iaccarino JM, Lange C, et al. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline [J]. *Eur Respir J*, 2020, 56(1): 200053. DOI: org/10.1183/ 13993003.00535-2020.
- [13] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis [J]. *European journal of cancer*, 2012, 48(4): 441-446. DOI:10.1016/j.ejca.2011.11.036

- [14]FALKINHAM J O 3RD. Nontuberculous mycobacteria in the environment[J]. Clin Chest Med.2002, 23(3): 529-551.
- [15]ABOAGYE S Y, DANSO E , ABOAGYE S Y, et al. Isolation of nontuberculous mycobacteria from the environment of Ghanaian communities where Buruli ulcer is endemic[J]. Appl Environ Microbiol, 2016, 82(14): 4320-4329.
- [16]CHEN C Y, CHEN H Y, CHOU C H, et al. Pulmonary infection caused by nontuberculous mycobacteria in a medical center in Taiwan, 2005-2008[J]. Diagn Microbiol Infect Dis. 2012, 72(1): 47-51.
- [17]VINNARD C, LONGWORTH S, MEZOCHOW A, et al. Deaths related to nontuberculous mycobacterial infection in the United States, 1999-2014[J]. Ann Am Thorac Soc, 2016, 13(11): 1951-1955.
- [18]RINGSHAUSEN F C, WAGNER D, DE ROUX A, et al. Prevalence of nontuberculous mycobacterial pulmonary disease, Germany, 2009-2014[J]. Emerg Infect Dis, 2016, 22(6): 1102-1105.
- [19]MARRAS T K, MENDELSON D, MARCHAND-AUSTIN, et al. Pulmonary nontuberculous mycobacterial disease, Ontario, Canada, 1998-2010[J]. Emerg Infect Dis, 2013, 19(11): 1889-1891.
- [20]ERASMUS JJ, MEADAMS HP, FARRELL MA, et al. Pulmonary nontuberculous mycobacterial infection; radiologic manifestations 1[J]. Radiographics. 1999, 19(6): 1487-1503.
- [21]Aksamit TR, Philley JV, Griffith DE. Nontuberculous mycobacterial (NTM) lung disease: the top ten essentials. Respir Med, 2014, 108(3): 417-425.
- [22]Yao Jingjiang, He Yaqiong, Zhang Yalin, Clinical and MSCT manifestations of pulmonary disease caused by nontuberculous mycobacterium[J]. Chinese Journal of Medical Imaging Technology. 2017, 33(3): 414-418.
- [23]Dai Jie, Shi Jingyun, Liang Li, et al. Imaging findings of nontuberculous mycobacterial lung diseases: a comparison with findings of secondary pulmonary tuberculosis[J]. The Journal of the Chinese Antituberculosis Association, 2014, 36(8): 706-709.
- [24]Albelda SM, Kern JA, Marinelli DL, et al. Expanding spectrum of pulmonary disease caused by nontuberculous mycobacteria. Radiology, 1985, 157: 289-296.
- [25]Moore EH. Atypical mycobacterial infection in the lung; CT appearance. Radiology, 1993, 187: 777-782.
- [26]Liu Y, Kim J, Balagurunathan Y, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas[J]. Clin Lung Cancer 2016; 17(5): 441-448. e6. DOI: 10.1016/j.clcc.2016.02.001

- [27] Ganeshan B, Panayiotou E, Burnand K, et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival[J]. *Eur Radiol* 2012;22(4):796-802.DOI:10.1007/s00330-011-2319-8
- [28] Gevaert O, Xu J, Hoang CD, et al. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results[J]. *Radiology* 2012 Aug;264(2):387-96.DOI:10.1148/radiol.12111607.
- [29]Zhang T1.Yuan M1.Zhong Y1.et al.Differentiation of focal organising pneumonia and peripheral adenocarcinoma in solid lung lesions using thin-section CT-based radiomics[J].*Clin Radiol.* 2019 Jan;74(1):78.e23-78.e30. DOI: 10.1016/j.crad.2018.08.014. Epub 2018 Oct 4.
- [30]Yanling W1,Duo G1,Zuojun G2,et al.Radiomics Nomogram Analyses for Differentiating Pneumonia and Acute Paraquat Lung Injury[J].*Sci Rep.*2019,Oct 21;9(1):15029. DOI: 10.1038/s41598-019-50886-7.
- [31]Zhang T1. Yuan M1. Zhong Y1. et al. Differentiation of focal organising pneumonia and peripheral adenocarcinoma in solid lung lesions using thin-section CT-based radiomics[J]. *Clin Radiol.* 2019 Jan; 74(1): 78.e23-78.e30.
- [32]Yanling W1. Duo G1. Zuojun G2. et al. Radiomics Nomogram Analyses for Differentiating Pneumonia and Acute Paraquat Lung Injury[J]. *Sci Rep.* 2019 Oct 21; 9(1): 15029.
- [33] Hayashi M,Takayanagi N, Kanauchi T,et al.Prognostic factors of 634 HIV-negative patients with Mycobacterium avium complex lung disease. *Am J Respir Crit Care Med*,2012,185(5):575–83.DOI:10.1164/rccm.201107-1203OC.
- [34] Gochi M,Takayanagi N,Kanauchi T,et al.Retrospective study of the predictors of mortality and radiographic deterioration in 782 patients with nodular/bronchiectatic Mycobacterium avium complex lung disease.*BMJ Open*2015;5:e008058.DOI:10.1136/bmjopen-2015-008058
- [35]Rosenzweig DY.Pulmonary mycobacterial infections due to Mycobacterium intracellulare-avium complex.Clinical features and course in 100 consecutive cases. *Chest*.1979;75:115-9.DOI:10.1378/chest.75.2.115
- [36]Lam PK,Griffith DE,Aksamit TR,et al.Factors related to response to intermittent treatment of Mycobacterium avium complex lung disease.*Am J Respir Crit Care Med.* 2006;173:1283-9.DOI:10.1164/rccm.200509-1531OC
- [37]Steven A,Cowman,Joseph Jacob,et al.Latent class analysis to define radiological subgroups in pulmonary nontuberculous mycobacterial disease[J].*BMC Pulmonary Medicine* 2018 Aug;18(1):145.DOI: 10.1186/s12890-018-0675-8

[38] Chen Lu, Shi Heshui, Li Benmei, et al. Comparative analysis of multi-slice spiral CT signs of nontuberculous mycobacterial lung disease and secondary tuberculosis [J]. Journal of Clinical Radiology, 2019, 38 (8):1400-1404.

Figures

Figure 1

A patient (male, 54 years old) with a cavity of NTM pulmonary disease in the right superior lobe. Outline along the outer edge of the cavity and include the cavity wall.

Figure 2

A patient (male, 65 years old) with a similar cavity of PTB in the right superior lobe. Outline along the outer edge of the cavity and include the cavity wall.

Figure 3

Variance threshold on feature select. We selected 857 radiomics features using variance threshold method (variance threshold = 0.8). The abscissa was the number of features, and the ordinate was the filtering method. Light blue in the bar indicates the number of original features, and dark blue indicated the number of filtered features.

Figure 5

Lasso algorithm on feature selectting. (a) Laso path; (b) MSE path; (c) coefficients in Lass model. Using Lasso model, 66 features which were correspond to the optimal alpha value were selected.

Figure 6

ROC curves of KNN methods to classification. (a) ROC curve of training set, the AUC were 0.988690981057 in NTM (sensitivity and specificity were 0.92 and {"TB": 0.92, "NTM": 0.96}), 0.988690981057 in TB (sensitivity and specificity were 0.96 and {"TB": 0.92, "NTM": 0.96}) respectively; (b) ROC curve of validation set, the AUC were 0.987012987013 in NTM (sensitivity and specificity were 0.94 and {"TB": 0.94, "NTM": 0.96}), 0.987012987013 in TB (sensitivity and specificity were 0.96 and {"TB": 0.94, "NTM": 0.96}) respectively.

Figure 7

ROC curves of SVM methods to classification. (a) ROC curve of training set, the AUC were 0.979926491377 in NTM (sensitivity and specificity were 0.95 and {"TB": 0.95, "NTM": 0.96}), 0.979926491377 in TB (sensitivity and specificity were 0.96 and {"TB": 0.95, "NTM": 0.96}) respectively; (b) ROC curve of validation set, the AUC were 0.983766233766 in NTM (sensitivity and specificity were 0.97 and {"TB": 0.97, "NTM": 0.82}), 0.983766233766 in TB (sensitivity and specificity were 0.82 and {"TB": 0.97, "NTM": 0.82}) respectively.

Figure 8

ROC curves of XGBoost methods to classification. (a) ROC curve of training set, the AUC were 0.999858637263 in NTM (sensitivity and specificity were 0.98 and {"TB": 0.98, "NTM": 0.99}), 0.999858637263 in TB (sensitivity and specificity were 0.99 and {"TB": 0.98, "NTM": 0.99}) respectively; (b) ROC curve of validation set, the AUC were 0.984848484848 in NTM (sensitivity and specificity were 0.97 and {"TB": 0.97, "NTM": 0.96}), 0.984848484848 in TB (sensitivity and specificity were 0.96 and {"TB": 0.97, "NTM": 0.96}) respectively.

Figure 9

ROC curves of RF methods to classification. (a) ROC curve of training set, the AUC were 0.999964659316 in NTM (sensitivity and specificity were 0.99 and {"TB": 0.99, "NTM": 1.0}), 0.999964659316 in TB (sensitivity and specificity were 1.00 and {"TB": 0.99, "NTM": 1.0}) respectively; (b) ROC curve of validation set, the AUC were 0.988636363636 in NTM (sensitivity and specificity were 0.94 and {"TB": 0.94, "NTM": 0.96}), 0.988636363636 in TB (sensitivity and specificity were 0.96 and {"TB": 0.94, "NTM": 0.96}) respectively.

Figure 10

ROC curves of LR methods to classification. (a) ROC curve of training set, the AUC were 0.986853265479 in NTM (sensitivity and specificity were 0.95 and {"TB": 0.95, "NTM": 0.95}), 0.986853265479 in TB (sensitivity and specificity were 0.95 and {"TB": 0.95, "NTM": 0.95}) respectively; (b) ROC curve of validation set, the AUC were 0.975108225108 in NTM (sensitivity and specificity were 0.94 and {"TB": 0.94, "NTM": 0.93}), 0.975108225108 in TB (sensitivity and specificity were 0.93 and {"TB": 0.94, "NTM": 0.93}) respectively.

Figure 11

ROC curves of DT methods to classification. (a) ROC curve of training set, the AUC were 1.0 in NTM (sensitivity and specificity were 1.00 and {"TB": 1.0, "NTM": 1.0}), 1.0 in TB (sensitivity and specificity were 1.00 and {"TB": 1.0, "NTM": 1.0}) respectively; (b) ROC curve of validation set, the AUC were 0.847402597403 in NTM (sensitivity and specificity were 0.91 and {"TB": 0.91, "NTM": 0.79}), 0.847402597403 in TB (sensitivity and specificity were 0.79 and {"TB": 0.91, "NTM": 0.79}) respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary.pdf](#)