The value of CT radiomics in differentiating mycoplasma pneumoniae pneumonia from streptococcus pneumoniae pneumonia with similar consolidation in children

Dongdong Wang (517555094@qq.con)
Qilu Hospital of Shandong University
Jianshe Zhao
Qilu Children's Hospital of Shandong University
Ran Zhang
Institute of Medical Technology
Qinghu Yan
Shandong Provincial Public Health Clinical Center: Shandong Provincial Chest Hospital
Lu Zhou
Qilu Hospital of Shandong University
Xiaoyu Han
Qilu Hospital of Shandong University
Yafei Qi
Qilu Hospital of Shandong University
Dexin YU
Qilu Hospital of Shandong University

Research Article

Keywords: CT, pneumonia, mycoplasma pneumoniae pneumonia, streptococcus pneumoniae pneumonia, radiomics

Posted Date: March 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1272026/v1

The value of CT radiomics in differentiating mycoplasma pneumoniae pneumonia from streptococcus pneumoniae pneumonia with similar consolidation in children

Dongdong Wang 1, Jianshe Zhao2,Ran Zhang3, Qinghu Yan4,Lu Zhou5,Xiaoyu Han1,Yafei Qi1,Dexin YU1 *

1 Department of Radiology, Qilu Hospital,Cheeloo College of Medicine,Shandong University, Jinan,Shandong,250012,China.

2 Department of Radiology, Qilu Children's Hospital, Jinan,Shandong,250012,China.

3 Huiying Medical Technology (Beijing) Co., Ltd. Beijing,100000,China

4 Department of Ultrasound, Shandong Public Health Clinical Center, Jinan,Shandong,250012,China.

5 Department of Cardiac Surgery ICU, Qilu Hospital，Cheeloo College of Medicine,Shandong University, Jinan,Shandong,250012,China.

Abstract

Objective To investigate the value of CT radiomics in the differentiation of mycoplasma pneumoniae pneumonia (MPP) from streptococcus pneumoniae pneumonia (SPP) with similar CT manifestations in children.

Methods 63 children with MPP (n=30) or SPP (n=33) with similar consolidation and surrounding halo on CT images in Qilu Hospital and Qilu Children's Hospital between January 2017 and July 2019 were enrolled in the retrospective study. Radiomic features of the both lesions on plain CT images were extracted including the consolidation part of the pneumonia or both consolidation and surrounding halo area which were respectively delineated at region of interest

(ROI) areas on the maximum axial image. The training set (n=43) and the validation set (n=20) were established by stratified random sampling at a ratio of 7:3. By means of variance threshold, the effective radiomics features, SelectBest and least absolute shrinkage and selection operator (LASSO) regression method were employed for feature selection. Six classifiers, including k-nearest neighbor(KNN), support vector machine (SVM),extreme gradient boosting(XGBoost), random forest(RF), logistic regression(LR), and decision tree(DT) were used to construct the models based on radiomic features.The diagnostic performance of these models was estimated and compared using the area under the receiver operating characteristic (ROC) curve (AUC), and the accuracy (score) matrix was established to compare and evaluate the results of different radiomics models.

Results XGBoost outperformed other classifiers and was selected as the backbone in the classifier with the consolidation + the surrounding halo was taken as ROI to differentiate MPP from SPP in validation set.The AUC value of MPP in validation set was 0.889, the sensitivity and specificity were 0.89 and 0.82, respectively;and the AUC value of SPP validation set was 0.889, the sensitivity and specificity were 0.82 and 0.89, respectively.

Conclusion The XGBoost model has the best classification efficiency in the identification of MPP from SPP in children, and the ROI with both consolidation and surrounding halo is most suitable for the delineation.

Key words: CT; pneumonia; mycoplasma pneumoniae pneumonia; streptococcus pneumoniae pneumonia; radiomics

Background

Community-acquired pneumonia is the leading cause of death among children under five years of age worldwide [1-3]. Mycoplasma pneumoniae pneumonia (MPP) and Streptococcus pneumoniae pneumonia (SPP) are common types of pneumonia in children[4, 5]. Early identification of the etiology and corresponding treatment can significantly reduce the mortality rate, however reliable samples of the biological causes of childhood pneumonia are difficult to obtain in clinical practice[6]. In addition, the colonization of pathogenic microorganisms in upper respiratory tract samples[7], nasal wipes, oropharyngeal wipes or sputum detection cannot accurately reflect the infection of the lower respiratory tract [8, 9]. At the same time, invasive lung puncture biospy, bronchoalveolar lavage, and other techniques cannot be used as routine detection methods in children. Currently, common pathogen detection methods have many defects such as long detection cycle time, false positive, and false negative results [10-13], which are greatly limited in practical clinical applications. Meanwhile, the imaging manifestations of MPP and SPP in most cases are similar in practical clinical situation, leading to difficulties in the differential diagnosis. Thus the timely and effective treatment requires accurate etiological details [14].

In recent years, the radiomics analysis based on massive data and artificial intelligence has shown significant advantages in judging disease types, predicting risk, and guiding treatment[15-17]. Radiomics converts medical images into high-dimensional images and mines effective data features through quantitative high-throughput extraction for data analysis, so various information that cannot be identified by the naked eye of radiologists, such as texture features, can be extracted , which is helpful for the diagnosis and treatment of diseases [18]. At the same time, this technology is simple and quick, and has a potential to solve the identification problems

between both the pneumonias. It is especially suitable for the patients who cannot obtain the results of pathogen detection in a short period of time, but are in critical condition and need accurate medication urgently.

We speculated that radiomics may be able to find more information that is not visible to the naked eye and may facilitate the differentiation of these two types of diseases. In this study, we collected a group of pediatric MPP and SPP patients with similar CT manifestations and difficulty in visual differentiation to investigate the value of CT radiomics in the differentiation of MPP and SPP. To the best of our knowledge, this is the first study to investigate the identification of pneumonia in children by radiomics.

METHODS

Study Design

This study was approved by the Ethics Committee of Qilu Hospital of Shandong University and Qilu Children's Hospital. Children with MPP and SPP who were admitted to Qilu Hospital of Shandong University or Qilu Children's Hospital on January 1, 2017 and July 1, 2019 were collected. All cases were diagnosed by clinical features and nucleic acid detection from bronchoalveolar lavage fluid. The chest CT images of the enrolled children were retrospectively analyzed. A total of 63 children were enrolled in this study, including 30 children with mycoplasma infection, with an average age of 23.6 ± 14.7 months. The mean age of 33 patients with Streptococcus pneumoniae was 19.8 ± 19.2 months. All the children were consistent with the clinical manifestations of pneumonia, such as fever, cough, and the presence of corresponding imaging findings. Then, 63 cases of pediatric pneumonia patients enrolled in the study were stratified and randomly sampled in a ratio of 7:3. All patients were divided into the training set (43

cases) and the validation set (20 cases).

Inclusion criteria: 1. Bronchoscopy was collected for alveolar lavage fluid for multiple PCR detection in all cases. Mycoplasma infection and Streptococcus pneumoniae infection were confirmed, and the corresponding treatment was effective; 2. CT images showed predominantly solid lesions on the lung window; 3. Two experienced associate professors in chest imaging diagnosis were unable to identify the nature of pneumonia. Exclusion criteria: 1. CT images have motion artifacts, poor image quality, large differences in scanning conditions, and inconsistent slice thickness; 2. There is clinical suspicion of mixed infection.

CT scanning method

SOMATOM Definition AS 64-slice spiral CT scanner was used to perform conventional chest scanning with the range from lung tip to the upper abdomen level of 5 cm below the diaphragmatic dome. The scanning parametric were as follows: tube voltage 120KV, tube current 250 ~ 400 mA/s (using automatic tube current modulation), FOV: 18 ~ 35 cm, matrix 512 × 512, slice thickness 5 mm, slice spacing 5 mm, scanning time 1.0 s. Patients who did not coordinate with the examination were routinely given sedative drugs.

Image delineation

The flow-chart depicting image of feature extraction and selection and and model construction is presented in Figure 1. The ROIs of the lesions on all lung window CT images were assessed and delineated in a double-blind manner by two radiologists with 5 and 10 years of experience respectively, and following review was performed by a senior physician. If the difference was ≥ 5%, the latter would determine the boundary and redraw it. Two different kinds of ROI delineation of the lesion were made: the first included the single consolidation part of the

lesion (Figure 2-b and d, blue line) and the second included the consolidation and surrounding halo area (Figure 2-b and d, orange line) . Meanwhile, the mediastinal window images was also used for the judgement of consolidation part of the lesion as the reference. The cavity, necrosis, hemorrhage or ground glassin oppacities in lesion were also included in the ROI. At the same time, the adjacent mediastinum, thickening pleura, and pleural effusion were avoided to draw by referring to these structures on mediastinal window CT. Grayscale normalization was then performed to reduce the impact of contrast and brightness changes. Ultimately, 63 ROIs were segmented from CT images of 63 patients and used for subject analysis.

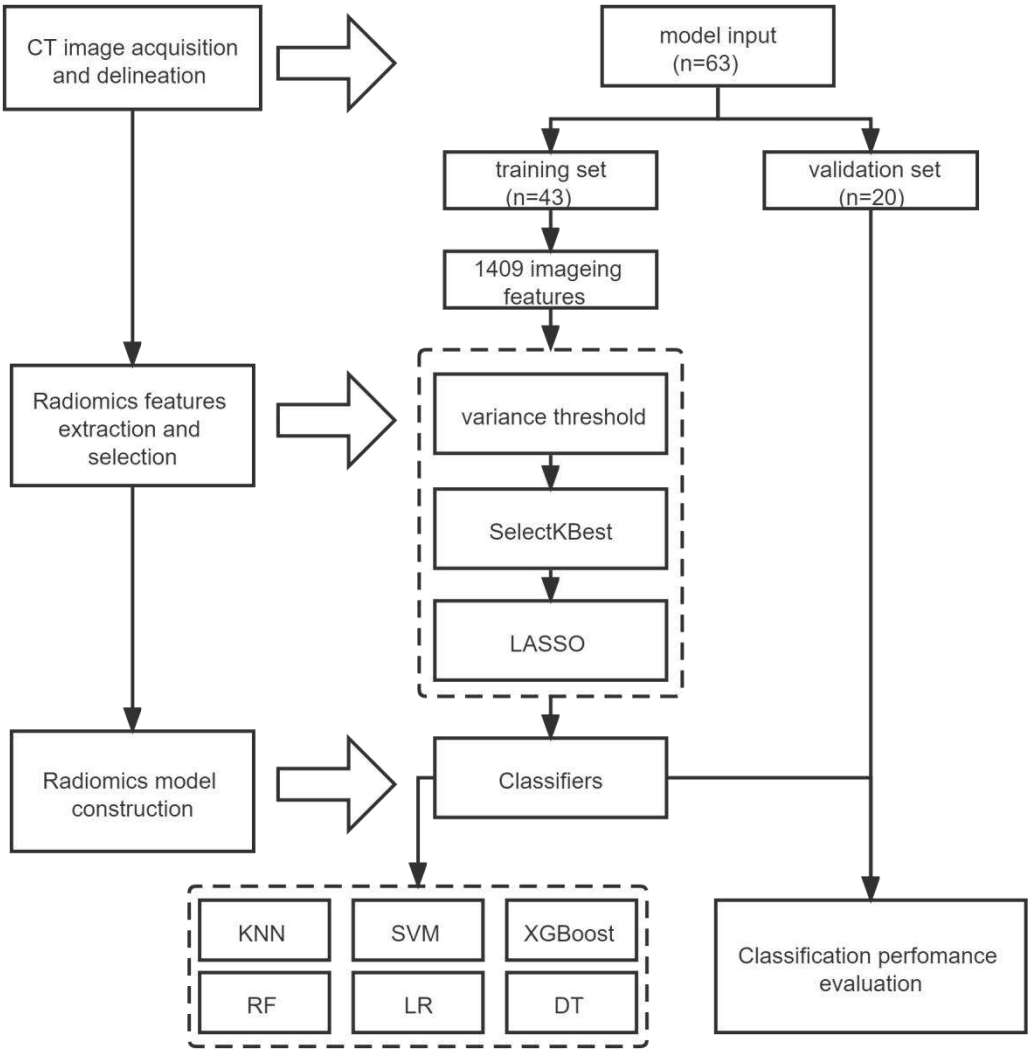

Figure 1.Fowchart of the whole radiomics study

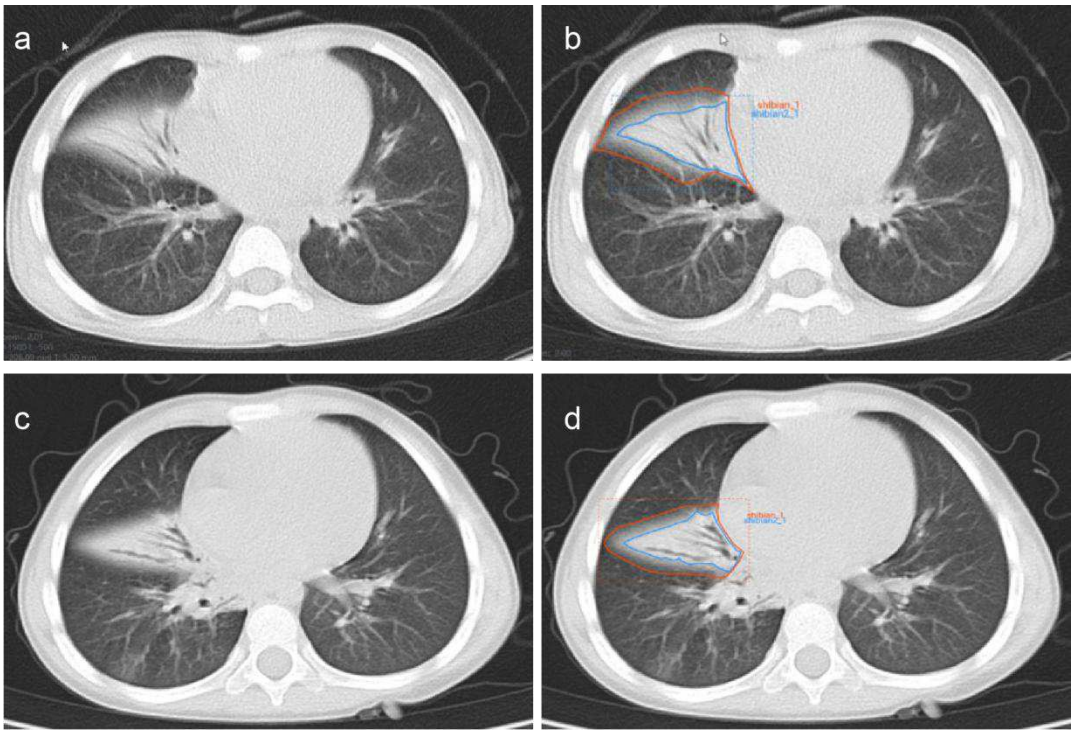

Figure 2.Manual delineation on lung window CT images in a 25-month female patient with mycoplasma pneumoniae pneumonia (Fig. 2.- a and b) and a 21-month male patient with streptococcus pneumoniae pneumonia (Fig. 2.- c and d). CT shows the similar appearances with consolidation and surrounding halo in middle lobe of right lung (a and c), and two ROIs (blue line and orange line) are delineated in each patient.

Feature extraction and selection

A total of 1409 quantitative image features were extracted from CT images using a RadCloud platform (http://radcloud.cn/). These properties were divided into three groups. The first group (first order statistics) consisted of 126 descriptors that could quantitatively describe the distribution of voxel intensities on CT images through commonly used basic indicators. The

second group (shape- and size-based features) reflecting the shape and size of the region. According to the operation length of gray level and the calculation of gray level co-occurrence texture matrix, 525 texture features that could quantify regional heterogeneity differences were classified into 3 categories (texture features). Texture features then characterized the recurrent local patterns on the image with their arrangement rules, including 75 features Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM) and Gray Level Size Zone Matrix (GLSZM). In addition, the texture was multi-resolution represented by filtering the image using 14 filters including index, logarithm, gradient, square value, square root, lbp-2D and wavelet (Wavelet-LHL, Wavelet-LHH, Wavelet-HLL, Wavelet-LLH, Wavelet-HLH, Wavelet-HHH, Wavelet-HHL, and Wavelet-LLL) to analyze the texture on a finer scale.

For intra-observer and inter-observer variation, intra-observer and inter-observer consistency of each feature was quantified by intra-class correlation (ICC) between calculated feature pairs, features with low reproducibility were excluded from subsequent analysis, and any feature with an ICC less than 0.8 was discarded. Redundant features could be reduced by reducing and selecting features to obtain the best results. The feature selection method used variance threshold (variance threshold = 0.8), SelectKBest and LASSO models. For the variance threshold method, the threshold was 0.8, so that the eight values of the variance smaller than 0.8 were removed. The SelectKBest method belonged to the univariate feature selection method, which used p-values to analyze the relationship between features and classification results, and all features with p-values less than 0.05 would be used. For the LASSO model, the L1 regularizer was used as the cost function, the error value of cross-validation was 5, and the maximum number of iterations was 1000.

Radiomics model construction

Based on the clinical data and the follow-up imaging analysis, the training and validation data sets were stratified and randomly sampled at a ratio of 7:3 to establish the training set (n=43) and the validation set (n=20), and the number of random seeds was 606. Six classifiers, including KNN, SVM, XGBoost, RF, LR and DT, were used to construct an Radiomics-based machine learning model to model MPP with SPP.

Statistical Analysis

SPSS statistics 26.0 software was used for statistical analysis of the data. Age difference in both diseases was tested by independent sample t test. Chi square test was used to analyze the differences in gender and inflammatory site. The linear combination of the selected features and the product of the corresponding weighting coefficients was used to form the radiomics label for each patient. ROC analysis was used to evaluate the diagnostic performances of the classifiers, and the accuracy (score) matrix was established to compare and evaluate different radiomics models. $P < 0.05$ was considered to indicate statistical significance.

Results

Study Population

There were no significant differences in age, gender or inflammatory site between MPP and SPP patients ($P>0.05$), as shown in Table 1.

Table 1 Comparison of patients' general information

Cases	Month	Gender		Location		
		Male	Female	Left lung	Right lung	Two lungs

MPP	30	23.6±14.7	15	15	6	14	10
SPP	33	19.8±19.2	22	11	10	13	10
P-value	-	0.248	0.180		0.639		

Feature extraction and screening results

In this study, the variance threshold method was used to screen out 449 features from 1209 CT features tested by ICC (Figure 3a), then 34 features were screened out by SelectKBest methods (Figure 3b), and finally 7 optimal features were screened out by LASSO algorithm (Figure 3c-e). The radiology score (Rad-score) formula was constructed based on these seven features and their regression coefficients (Table 2), and the formula was: Rad-score = feature * coefficient.

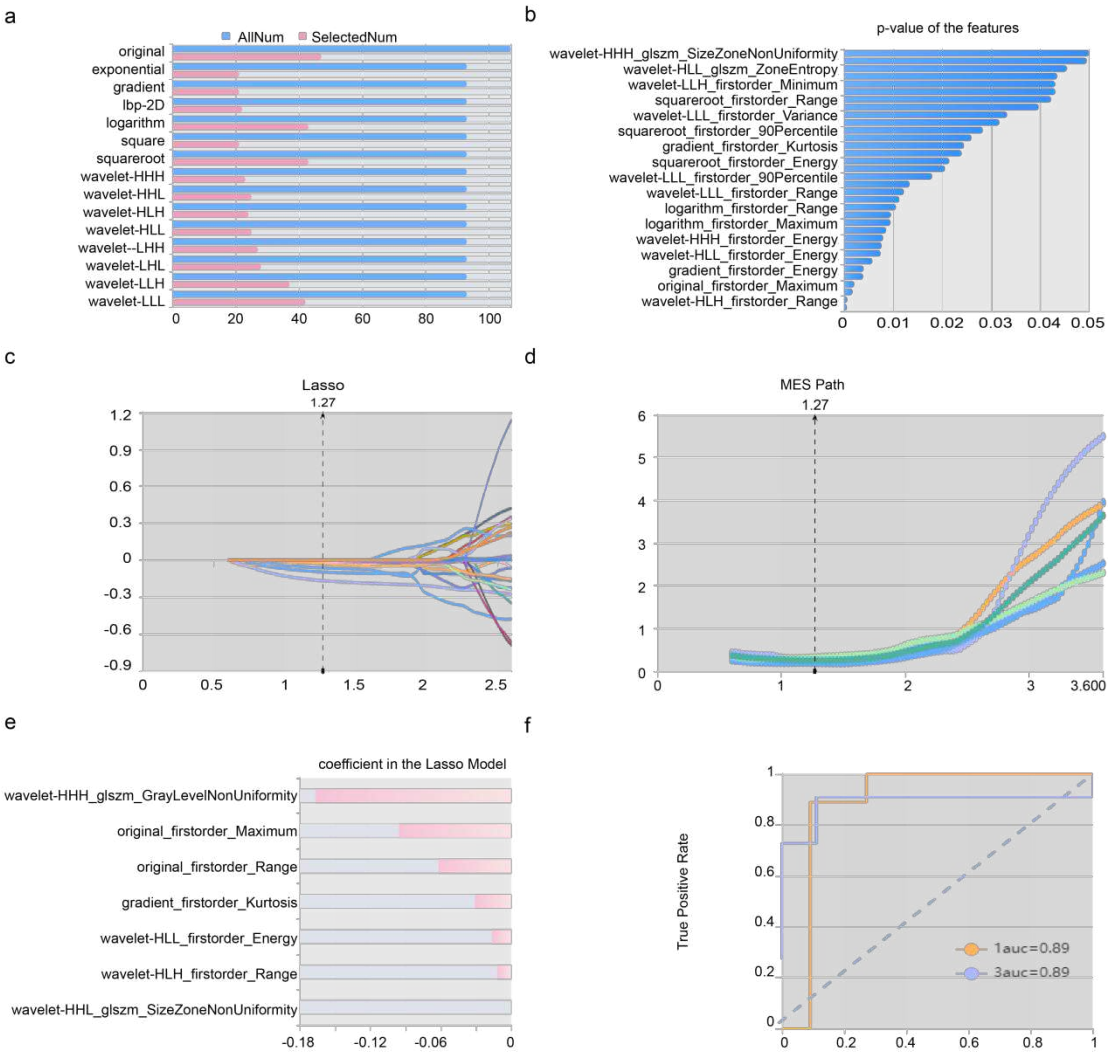

Figure 3. Workflow model construction and radiomics analysis. Fig.3-a is a Variance threshold on feature select.The blue bar represents the number of all the extracted radiomics features, and the pink bar represents the number of radiomics features screened by variance threshold method. The vertical axis is 15 kinds of filtering methods (variance threshold =0.8). Fig.3-a is Select K best on feature select. The abscissa is the P value of the feature, and the ordinate is the feature whose P value < 0.05 is screened by SleckBest method.Fig.3-c-e are schematic diagram of feature screening by Lasso method: c Lasso path, where the abscissa is the log value of α, and the ordinate is the coefficient of the feature;d The abscissa of the MSE path is the log value of α, and the ordinate is the mean square error;e Regression coefficient of Lasso model, where the abscissa

represents the regression coefficient and the ordinate represents the selected features.f ROC curve of the XGBoost model.Yellow curve is MPP set, blue curve is SPP set.

Table 2 Radiomics features and their categories, filters and regression coefficients selected with ROI in the consolidation part and surrounding halo area

Radiomic feature	Radiomic class	Filter	Coefficient
Range	firstorder	wavelet-HLH	- 0.01227
GrayLevelNonUniformity	glszm	wavelet-HHH	- 0.16672
Maximum	firstorder	original	- 0.09587
Range	firstorder	original	- 0.06199
Energy	firstorder	wavelet-HLL	- 0.01700
Kurtosis	firstorder	gradient	- 0.03129
SizeZoneNonUniformity	glszm	wavelet-HHL	- 0.00042

Differential efficacy of radiomics models

The analysis results of accuracy (score) matrix in the training and validationt sets of six models with two kinds of ROI is shown in Table 3.All classifiers in consolidation region are shown in Table 4. After the ROI contained surrounding halo area, the matrix scores of all classifiers were significantly improved. The ROC curve analysis results of all classifier in validation sets are shown in Table 5 and the XGBoost classifier get the best matrix score. When this classifier was used for training set, the AUC of MPP in validation set was 0.889 (95% confident interval [CI]: 0.72 ~ 1.00, sensitivity= 0.89, specificity= 0.82), and the AUC of SPP in validation set was 0.889 (95% CI:0.72~1.00, sensitivity=0.82, specificity=0.89) (Figure 3f). The four indicators of the classifier (accuracy, recall rate, F1 score, and support) are shown in Table 6.

Table 3 Accuracy (score) matrix in the training and validationt sets of six models with two kinds of ROI

	Validation_score of the consolidation area	Training_score of the consolidation area	Validation_score of the consolidation+surrounding halo	Training_score of the consolidation+surrounding halo
KNN	0.68	0.77	0.70	0.8
SVM	0.58	0.77	0.75	0.91
XGBoost	0.63	0.95	0.8	0.93
RF	0.63	0.95	0.6	1
LR	0.58	0.74	0.75	0.86
DT	0.53	1	0.7	1

Table 4 ROC curve analysis results in validation set with ROI of the consolidation region

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	MPP	0.628	0.41 - 0.84	0.56	0.80
	SPP	0.628	0.41 - 0.84	0.80	0.56
SVM	MPP	0.700	0.47 - 0.93	0.67	0.60
	SPP	0.700	0.47 - 0.93	0.60	0.67
XGBoost	MPP	0.700	0.48 - 0.92	0.67	0.70
	SPP	0.700	0.48 - 0.92	0.70	0.67
RF	MPP	0.650	0.44 - 0.86	0.33	0.70
	SPP	0.650	0.44 - 0.86	0.70	0.33
LR	MPP	0.689	0.45 - 0.92	0.67	0.50
	SPP	0.689	0.45 - 0.92	0.50	0.67
DT	MPP	0.422	0.19 - 0.66	0.44	0.40
	SPP	0.422	0.19 - 0.66	0.40	0.44

Table 5 ROC curve analysis results in validation set with ROI of both the consolidation part and surrounding halo area

Classifiers	Category	AUC	95% CI	Sensitivity	Specificity
KNN	MPP	0.788	0.58 - 1.00	0.78	0.64

	SPP	0.788	0.58 - 1.00	0.64	0.78
	MPP	0.747	0.54 - 0.95	0.78	0.73
SVM					
	SPP	0.747	0.54 - 0.95	0.73	0.78
	MPP	0.889	0.72 - 1.00	0.89	0.82
XGBoost					
	SPP	0.889	0.72 - 1.00	0.82	0.89
	MPP	0.783	0.55 - 1.00	0.67	0.55
RF					
	SPP	0.783	0.55 - 1.00	0.55	0.67
	MPP	0.828	0.62 - 1.00	0.78	0.73
LR					
	SPP	0.828	0.62 - 1.00	0.73	0.78
	MPP	0.697	0.48 - 0.92	0.67	0.73
DT					
	SPP	0.697	0.48 - 0.92	0.73	0.67

Table 6. Evaluation results of the four indicators of the both diseases in validation set with ROI of both the consolidation part and surrounding halo area

	Indicators	KNN	SVM	XGBoost	RF	LR	DT
MPP	Precision	0.64	0.70	0.80	0.55	0.70	0.64
	Recall	0.78	0.78	0.89	0.67	0.78	0.78
	F1-score	0.70	0.74	0.84	0.60	0.74	0.70
	Support	9	9	9	9	9	9
SPP	Precision	0.78	0.80	0.90	0.67	0.80	0.78
	Recall	0.64	0.73	0.82	0.55	0.73	0.64
	F1-score	0.70	0.76	0.86	0.60	0.76	0.70
	Support	11	11	11	11	11	11

Discussion

In this retrospective study,we established six classifiers to evaluated the performance for discriminating MPP from SPP with two kinds of ROI.We found that the radiomics classifer demonstrated low performance for differentiation using the ROI of consolidation region. After the ROI contained surrounding halo area,,every classifiers performed better and demonstrated high performance.Our preliminary results revealed that classifiers trained with ROI of both the consolidation part and surrounding halo area achieved better diagnostic performance for discrimination between MPP and MPP,with significantly higher AUC than the ROI of consolidation region in the validation set and in all patients.

The differentiation between MPP and SPP in children is pivotal, as treatment approaches are quite different[19]. Accurate and early etiological diagnosis can guide clinicians in rational drug

using[20], reducing the total mortality rate by 27% and the pneumonia-specific mortality rate by 42%[21]. At the same time, it can also reduce the abuse of antibiotics and improve bacterial drug resistance. Therefore, the accurate diagnosis and identification is helpful in guiding treatment.

In radiomics, there are a variety of machine learning methods that can be used to build classification models, which have their own advantages for different tasks. In this study, six commonly used classifier models (KNN, SVM, XGBoost, RF, LR, and DT) and 1,409 features for each patient were extracted, and then seven features were screened out by LASSO algorithm finally. Among the seven significant radiomics features, we found the mean intensity (texture feature) of MPP was different from SPP. This might reflect different diffuse opacities or greater degree of fluid or debris affecting the airspaces leading to the diversity in airspace disease phenotypes (consolidation, ground-glass opacities, etc.) that combine varying degrees of edema and vascular and interlobular septal thickening[22, 23],so the MPP showed different irregular intensity changes, heterogeneous intensities, and range in textures from SPP. When the inflammation progresses, extensive exudation of lung tissue leads to solid pneumonia, and the early specific imaging findings of two types of inflammation were obscurated. This might also explain why the consolidation part of the pneumonia was limited in differentiation.

To date, very few studies have addressed the problem of pneumonia differentiation using radiomics. Mei et al.[24] used artificial intelligence algorithms to integrate chest CT findings with clinical symptoms, exposure history and laboratory validation to diagnose COVID-19. Wang et al.[25]combined deep learning-radiomics model to differentiate COVID-19 from non-COVID-19 viral pneumonia. These studies demonstrate the feasibility of using radiomics to identify lung inflammation.However, their study analysis showed only one kinds of pneumonia, and they did

not find a significant effect of different ROI delineation methods on the results. Our approach is more pragmatic, as using only one kinds of pneumonia may introduce a selection bias and overestimate the classification accuracy. Yu et al. [26]implemented multiple network architectures to subclassify NSCLCand achieved an AUC of 0.864, which was 0.025 lower than our result. Further more, Wang et al. [27] conducted a similar classification task their CNN-AvgFea-Norm3-based RF method achieved an AUC of 0.856 and an accuracy of 0.820, which was 0.069 loss in AUC and 0.070 loss in accuracy compared with our classifier. This may be related that our ROI includes more information

XGBoost radiomics classifer is used widely by scientists for solving realworld scale problems with limited resources[28-31]. The building model and following validition results proved that XGBoost radiomics classifer showed a good value in differentiation of mycoplasma from pneumococcal pneumonia in children, consistent with their different pathological basis. The type of pathogens in pneumonia are associated with different lymphocytes and monocytes response during the initiation[32] and hinted that radiomic model could be used as the identification method for multiple kinds of pneumonia. This provides clinicians with additional diagnostic information and promote the development of personalized precision therapy.

Although we found high diagnostic performance of radiomics model, the study still has some limitations. First, the number of patients is not large enough.,External validation cannot be done due to insufcient data and the diagnostic accuracy might be overestimated.Hence,large multicenter studies are warranted to confirm the current findings.Secondly,the study only performed two-dimensional analysis of the largest section,it may lead to inaccurate evaluation due to the

influence of acquisition parameters on feature stability. Finally,the study only discuss two comment types of pneumonia in children and the others were not included.

Conclusion

In conclusion,The XGBoost radiomics classifer with the ROI delineation including both solidation and surroungding halomay show potential for the differentiation of MPP and SPP, which may provide accurate differentiational diagnosis for the early and appropriate treatment.

Declarations:

Ethics approval and consent to participate

The study was conducted in accordance with the principles of the Declaration of Helsinki,and the study protocol was approved by the ethics committee of Qilu Hospital with the number of 2021224.

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.

Financial support.

None

Authors' contributions

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

Acknowledgements

Thanks to Dr. Zhang for the polish.

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. Naucler, P., et al., *The changing epidemiology of community-acquired pneumonia: nationwide register-based study in Sweden.* J Intern Med, 2019. **286**(6): p. 689-701.
2. Wang, H., et al., *A versatile loop-mediated isothermal amplification microchip platform for Streptococcus pneumoniae and Mycoplasma pneumoniae testing at the point of care.* 2019. **126**: p. 373-380.
3. Liu, L., et al., *Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis.* 2015. **385**(9966): p. 430-40.
4. Chen, Y., W. Hsu, and T.J.E.i.d. Chang, *Macrolide-Resistant Mycoplasma pneumoniae Infections in Pediatric Community-Acquired Pneumonia.* 2020. **26**(7): p. 1382-1391.
5. Asner, S., et al., *Burden of Streptococcus pneumoniae Sepsis in Children After Introduction of*

Pneumococcal Conjugate Vaccines: A Prospective Population-based Cohort Study. 2019. **69**(9): p. 1574-1580.

6. Gray, D., et al., *Lung Function in African Infants in the Drakenstein Child Health Study. Impact of Lower Respiratory Tract Illness.* 2017. **195**(2): p. 212-220.

7. Turner, G., et al., *The role of postmortem studies in pneumonia etiology research.* 2012: p. S165-71.

8. Abanses, J., et al., *Impact of rapid influenza testing at triage on management of febrile infants and young children.* 2006. **22**(3): p. 145-9.

9. Charlson, E.S., et al., *Topographical continuity of bacterial populations in the healthy human respiratory tract.* Am J Respir Crit Care Med, 2011. **184**(8): p. 957-63.

10. Bénet, T., et al., *Microorganisms Associated With Pneumonia in Children <5 Years of Age in Developing and Emerging Countries: The GABRIEL Pneumonia Multicenter, Prospective, Case-Control Study.* 2017. **65**(4): p. 604-612.

11. Totten, A., et al., *Allergic airway sensitization impairs antibacterial IgG antibody responses during bacterial respiratory tract infections.* 2019. **143**(3): p. 1183-1197.e7.

12. Wang, L., et al., *A comparison study between GeXP-based multiplex-PCR and serology assay for Mycoplasma pneumoniae detection in children with community acquired pneumonia.* 2017. **17**(1): p. 518.

13. Rivaya, B., et al., *Macrolide resistance and molecular typing of Mycoplasma pneumoniae infections during a 4 year period in Spain.* 2020. **75**(10): p. 2752-2759.

14. Christensen, M., et al., *Simulated Adoption of 2019 Community-Acquired Pneumonia Guidelines Across 114 Veterans Affairs Medical Centers: Estimated Impact on Culturing and*

Antibiotic Selection in Hospitalized Patients. 2021. **72**: p. S59-S67.

15. Chaddad, A., et al., *Radiomics in Glioblastoma: Current Status and Challenges Facing Clinical Implementation.* 2019. **9**: p. 374.

16. Tobaly, D., et al., *CT-Based Radiomics Analysis to Predict Malignancy in Patients with Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas.* 2020. **12**(11).

17. Hsu, J., et al., *Radiomic Immunophenotyping of GSEA-Assessed Immunophenotypes of Glioblastoma and Its Implications for Prognosis: A Feasibility Study.* 2020. **12**(10).

18. Park, V.J.E., *Expanding applications of MRI-based radiomics in HER2-positive breast cancer.* 2020. **61**: p. 103085.

19. Conrad, D., et al., *Cystic fibrosis therapy: a community ecology perspective.* 2013. **48**(2): p. 150-6.

20. Mahony, J.J.C.m.r., *Detection of respiratory viruses by molecular methods.* 2008. **21**(4): p. 716-47.

21. Sazawal, S., R. Black, and J.T.L.I.d. , *Effect of pneumonia case management on mortality in neonates, infants, and preschool children: a meta-analysis of community-based trials.* 2003. **3**(9): p. 547-56.

22. Burgoyne, R.A., A.J. Fisher, and L.A. Borthwick, *The Role of Epithelial Damage in the Pulmonary Immune Response.* Cells, 2021. **10**(10).

23. Martin, F.P., et al., *Alveolar Macrophages: Adaptation to Their Anatomic Niche during and after Inflammation.* Cells, 2021. **10**(10).

24. Mei, X., et al., *Artificial intelligence-enabled rapid diagnosis of patients with COVID-19.* 2020. **26**(8): p. 1224-1228.

25. Wang, H., et al., *Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures.* 2021. **48**(5): p. 1478-1486.

26. Yu, K., et al., *Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks.* 2020. **27**(5): p. 757-769.

27. Wang, X., et al., *Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis.* 2020. **50**(9): p. 3950-3962.

28. Chen, X., Z. Wang, and X.J.S.r. Pan, *HIV-1 tropism prediction by the XGboost and HMM methods.* 2019. **9**(1): p. 9997.

29. Torlay, L., et al., *Machine learning-XGBoost analysis of language networks to classify patients with epilepsy.* 2017. **4**(3): p. 159-169.

30. Ogunleye, A., Q.J.I.A.t.o.c.b. Wang, and bioinformatics, *XGBoost Model for Chronic Kidney Disease Diagnosis.* 2020. **17**(6): p. 2131-2140.

31. Yu, D., et al., *Copy number variation in plasma as a tool for lung cancer prediction using Extreme Gradient Boosting (XGBoost) classifier.* 2020. **11**(1): p. 95-102.

32. Kreutmair, S., et al., *Distinct immunological signatures discriminate severe COVID-19 from non-SARS-CoV-2-driven critical pneumonia.* 2021. **54**(7): p. 1578-1593.e5.