The Value of CT Radiomics Nomogram Analysis of Parenchyma Surrounding Pulmonary Nodules in the Differentiation of Inflammatory Nodules from Tuberculous Nodules

WeiJie Fan
Xinqiao Hospital
Ke Mu
Xinqiao Hospital
Si Zhang
Xinqiao Hospital
Yan Yang
Xinqiao Hospital
Li Yu
Xinqiao Hospital
WenJie Huang
Xinqiao Hospital
Li Wen
Xinqiao Hospital
Huan Liu
GE Healthcare
Dong Zhang (zhangdongxqyy@163.com)
Xinqiao Hospital

Research

Keywords: Pulmonary nodules, Radiomics, Diagnosis, Nomogram

Posted Date: September 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-839717/v1

Abstract

Background: To investigate the value of CT Radiomics nomogram based on pulmonary nodules and surrounding parenchyma in differentiating non-tuberculous inflammatory pulmonary nodules from tuberculous nodules.

Methods: A retrospective analysis was performed on 273 patients with pulmonarynodules confirmed by surgery and pathology in the Second Affiliated Hospital of Army Military Medical University from January 2015 to March 2021, including 164 cases of non-tuberculous inflammatory nodules and 109 cases of tuberculous nodules.Pulmonary nodule (ROI1), 3mm parenchymal band around nodule (ROI2) and nodules with external expansion (ROI3) were segmented andradiomic features were extracted by A.K. software. The stability of the features was analyzed by ICC, and the features were divided into the training set and the verification set by random stratified sampling according to the ratio of 7:3. The radiomics label was constructed by using the maximum relevance and minimum redundancy method (mRMR) and the least absolute shrinkage and selection operator method (LASSO) after two dimensional reduction.Finally, multiple Logistic regression analysis was conducted to select the optimal model among the three models, and the classification model including the radiomicfeatures and clinical risk factors was established.The identification effectiveness of the model was evaluated in terms of average area under curve (AUC), accuracy, sensitivity, and specificity, and was independently validated in a validation cohort.

Results: There was significant difference in the type of pulmonary nodules between the two groups ($p<0.05$). Nodules with external expansion model has the best diagnostic performance, and combined with clinical features (nodular type) to construct a classification model.The AUC of both the training cohort and the validation cohort was 0.93. Decision curve analysis showed that the radiomics nomogram had good value in classification of the two categories of nodules.

Conclusions: The CT radiomics nomogram based on nodule plus surrounding substance has good performance in judging non-tuberculous inflammatory nodules and tuberculous nodules, which can provide reference for clinical diagnosis and treatment, and has good clinical application value and development prospect.

1. Background

The wide application of low-dose chest CT and artificial intelligence assisted diagnosis of pulmonary nodules has significantly improved the detection rate of pulmonary nodules[1, 2]. The nature of detected nodules can be either benign lesions such as inflammation and benign tumors, or malignant lesions such as lung cancer and metastatic tumors. Therefore, that can also cause a serious impact on the accurate diagnosis of clinicians and the psychological pressure of patients[3]. Due to the different nature of pulmonary nodules, their treatment options are different: malignant nodules usually require surgical resection; non-tuberculous inflammatory nodules require routine anti-inflammatory treatment; tuberculous

nodules should be treated with anti-tuberculous drugs. Therefore, precise diagnosis of pulmonary nodules can effectively guide clinical treatment and avoid the waste of medical resources caused by excessive medical treatment. Previous studies mostly focused on the diagnosis of benign and malignant pulmonary nodules, while there were few studies on the differentiation between non-tuberculous pulmonary nodules and tuberculous nodules. As a special type of inflammatory nodules, tuberculous nodules have different pathological basis from non-tuberculous inflammatory nodules[4]. Although, these two types of pulmonary nodules are difficult to distinguish by conventional CT diagnosis, but radiomics can be used to identify them by extracting information about the nodules and surrounding parenchyma. Radiomics is an emerging technology based on big data technology combined with computer-aided diagnosis (CAD)[5]. By mining and extracting a large amount of information invisible to the naked eye but of clinical value from medical images, it can provide pathological classification or predict the prognosis of diseases. In this study, two types of pulmonary nodules and their surrounding parenchyma were extracted with high-throughput features by Radiomics. Through the establishment of multiple classification models and the selection of the best diagnostic efficiency of the model to build the radiomic nomogram, which in order to improve the accuracy of pneumonia nodal classification diagnosis, and provide the basis for clinical precision treatment.

2. Materials And Methods:

2.1 Subject

This prospective study was approved by the Medical Ethics Committee of Second Affiliated Hospital of Army Medical University (2020-147-01). All subjects were exempted from informed consent, Moreover, all methods were implemented in accordance with the approved regulations and the Declaration of Helsinki. Clinical data and Lung Low-dose computed tomography (LDCT) imaging data of patients undergoing pulmonary nodules surgery from January 2015 to March 2021 were retrospectively collected. **Inclusion criteria**: (1) Patients were proved to be non-tuberculous inflammatory nodules or tuberculous nodules by operation and pathology. (2) the Lung LDCT was performed preoperatively, and showed no significant signs of pneumonia or other obvious signs of tuberculosis. (3) No invasive procedures such as puncture were performed before LDCT examination. (4) Pulmonary nodules have no radiographic features such as calcification and cavity of typical benign nodules. **Exclusion criteria**: (1) LDCT was not performed in our hospital preoperative. (2) LDCT images showed other lesions or artifacts such as respiratory movements. (3) Patients with a previous history of tuberculosis. (4) Patients who received treatment such as radiation or chemotherapy prior to LDCT examination. During this period, a total of 2,768 patients underwent surgery for pulmonary nodules, among which 2,323 patients were pathologically confirmed to be malignant pulmonary nodules, and the rest were benign nodules, including tuberculosis, chronic inflammation, cryptococcosis and pulmonary hamartoma. Inflammatory pulmonary nodules account for a large proportion of surgical treatment. In the end, 109 cases of tuberculous nodules and 164 cases of non-tuberculous inflammatory nodules were enrolled in this study.

Preoperative clinic data and laboratory data of patients were collected on hospital HIS system. General clinical data include gender and age; Inflammatory markers include white blood cell count, NEUT%, LYM%, MXD%, EO%, BASO%, C-reactive protein and erythrocyte sedimentation rate; Indicators of TB infection include PPD test and T-spot test; Tumor markers include CEA, CA153, NSE, SCC, cyfra21-1, SF, CA125 and ProGRP. The nodule type information of patients was collected by Picture Archiving and Communication Systems (PACS). Types of nodules include pure ground glass nodules (pGGN), mixed ground glass nodules (mGGN), and solid nodules (SN). This was determined with reference to the data annotation of pulmonary nodules and the consensus of quality control experts[6, 7].

2.2 Data acquisition

All subjects underwent lung LDCT examination using GE Optima CT660 (GE Healthcare, USA), LDCT Scanning parameters are as follows: tube voltage:100Kv; automatic tube current modulation; field of view: 500mm; detector collimation: 0.625mm; layer thickness and spacing: 0.625mm; gathering matrix: 512×512; pitch: 1.375. All imaging data were reconstructed without interval by using high resolution and standard lung algorithms, and the thickness of the reconstruction layer was 5mm. All clinical data of CT images were desensitized and exported in DICOM format by PACS.

2.3 Image preprocessing and ROI segmentation

The axial LDCT images with 0.625mm layer thickness of each subject were imported into the A.K (Artificial Intelligent Kit, version 3.3.0, GE Healthcare) research platform. The original image was preprocessed as follows: Firstly, the spatial resolution of the image was adjusted to 1 mm×1 mm × 1 mm by resampling the CT image. Then the CT images were standardization with the grayscale unified adjusted to 0 ~ 255. The final image after preprocessing, reference lung nodules data labeling and quality control expert consensus[6], by two radiologists (7 years and 9 years of experience in chest CT imaging diagnostic) for nodules semiautomatic segmentation and manual correction nodule boundary, by labeling leader and arbitration experts (both for More than 15 years of experience in diagnosis of senior professional doctor) for review and modification. The ROIs of pulmonary nodules was finally determined. ROI1 was segmented the pulmonary nodules, ROI2 was expanded by 3mm on the basis of nodular segmentation, ROI3 was the total of pulmonary nodules plus the external (Fig. 1A-F).

2.4 Feature extraction

Firstly, we imported all the two groups of preprocessed axial LDCT images. Then import all corresponding three ROIs in batches. We selected total 851 radiomics features of each segmented pulmonary nodules, which were strictly followed by Image Biomarker Standardization Initiative (IBSI)[8], Including the first order features, shape features, Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Size Zone Matrix (GLSZM), Gray-Level Run Length Matrix (GLRLM), Neigbouring GrayTone Difference Matrix (NGTIM), Gray Level Dependence Matrix (GLDM), and wavelet features. Wavelet transform is a local transform of time and frequency. It has the feature of multi-resolution analysis and can display the local characteristics of the signal in time domain and frequency domain[9]. The results of two-dimensional

wavelet decomposition reflect the frequency changes in different directions and the texture features of the image. All features were extracted using A.K software.

2.5 Feature selection

Before the feature selection, Intraclass Correlation Coefficient (ICC) was used to check the consistency of the radiomic features extracted by two surveyors to ensure the stability of the extracted radiomic features. ICC average > 0.75 was selected for subsequent analysis[10]. Then all subjects were divided into training set and verification set by random stratified sampling at a ratio of 7:3. Finally, the unit limitation for each column of features was eliminated by normalization. Two feature selection methods to select the radiomic features among three models of each ROI in two groups. At first, the maximum correlation and minimum redundancy (mRMR) method was used to reduce the dimension of stable radiomic features to eliminate redundant and irrelevant features, that is, to maximize the correlation between features and classification variables. Then, feature dimensionality is reduced again by Least Absolute Shrinkage and Selection Operator (LASSO), and the best parameters are obtained through cross verification. Lambda is used to carry out final feature Selection based on the principle of minimum error. At the same time, 10-fold cross validation is carried out to select the optimal subset of the best features to build the final three models.

2.6 Comprehensive radiomics signature construction and validation

Logistic regression analysis was performed on the selected radiomics features to establish the classification and diagnosis model for non-tuberculous inflammatory nodules and tuberculous nodules by ROI1, ROI2 and ROI3, respectively. The test data was analyzed by ROC to independently validate the performance of each diagnostic model, then choose the best diagnostic model. A Nomogram was established using the RMS software package in R language to evaluate the risk of pulmonary nodules being tuberculous. Based on the decision curve, the corresponding net benefit was calculated under different threshold probabilities to evaluate its value in clinical application, and the model fitting degree was tested by Hosmer-Lemeshow[11].

2.7 Statistical analysis

SPSS (version 22.0, IBM) was used for statistical analysis of the general data and laboratory indicators of the two groups of patients. Qualitative data were expressed as frequency and $\chi 2$ test was used. For the measurement data following normal distribution, the mean ± standard deviation was used to express, and the independent sample t test was used. For measurement data that do not obey the normal distribution, the median (upper and lower quaternary) is used to represent, and the non-parametric Mann-Whitney U test is used. In addition, image omics feature selection, ROC curve drawing, model construction and verification were all performed using R software (version 3.6.0; http://www.Rproject.org); All statistics were double-tailed analysis, and $P < 0.05$ was considered statistically significant.

3. Results

3.1 Demographic and neurocognitive

Clinical data of 263 cases of patients with general description as shown in table 1, The nodule type between two groups was statistically significant difference ($p < 0.05$). There was no difference between general data and laboratory data in the other two groups ($P > 0.05$). The tumor markers between the two groups were no different and they were not listed because the study was aim to analyze the benign pulmonary nodules. Due to the laboratory data on CRP, erythrocyte sedimentation rate, and tuberculosis indicators were not tested in most patients, comparative analysis was not performed.

Table 1 General clinical data and laboratory examination data

		Non-tuberculous inflammatory nodules	Tuberculous nodules	statistic	*P* value
Subjects（n）		164	109	NA	NA
Age（year）		50.31±9.39	48.39±10.47	0.908	0.342
Gender	male	94	70	NA	0.254#
	female	70	39		
Nodule size（mm）		12.02±4.67	13.85±4.25	0.062	0.803
Nodule type	pGGN	77	14	NA	0.000#
	mGGN	11	7		
	SN	76	88		
WBC（10^9/L）		5.89±1.90	5.79±1.79	0.079	0.779
NEUT%		59.24±9.85	61.98±9.50	0.095	0.760
LYM%		28.34±8.46	29.96±8.60	0.255	0.614
MXD%		6.95(6.10,8.30)	7.00 (5.40,8.10)	0.881	0.379*
EO%		2.05(1.30,3.18)	1.70(1.10,2.70)	1.788*	0.074*
BASO%		0.40(0.21,0.68)	0.40(0.27,0.60)	0.593*	0.553*

χ2 test was used for gender and nodule type;

* Nonparametric Mann-Whitney U test was used for data that do not follow normal distribution;

Two-sample *t* test was used for the normal distribution data.

3.2 Feature selection results

The 790 stable features retained by ICC, then each of three ROI models were remained with 30 radiomic features reduced by mRMR. Finally, LASSO regression analysis and 10-fold cross validation were used for screening, then the model ROI1, ROI2 and ROI3 were left with 9, 10 and 11 features respectively (Fig. 2A-C). After the number of feature determined, RadScores of training cohort and test cohort were calculated in each model. the most predictive subset of feature was chosen and the corresponding coefficients were evaluated (Fig. 3A-C).

3.3 Comprehensive radiomics signature construction and validation

The comparative analysis showed that there were significant differences in the RadScore distribution from class 0 and Class 1 of the three models on the training group and testing group respectively ($P < 0.05$) (Fig. 4A-C). Logistic regression analysis of patients' radiomics features was performed to obtain ROC curves to classify the two types of nodules, and the test data were used for internal independent verification in each model. Verification results of classification of non-tuberculous inflammatory pulmonary nodules and tuberculous nodules: Nodules model (ROI1) achieves an area under the ROC curve (AUC) of 0.91 (95% confidence interval 0.85 ~ 0.97) in the testing group, and have an accuracy of 0.838, sensitivity of 0.914, and specificity of 0.783. Nodules external model (ROI2) achieves an area under the ROC curve (AUC) of 0.91 (95% confidence interval 0.83 ~ 0.98) in the testing group, and have an accuracy of 0.846, sensitivity of 0.794 and specificity of 0.918. Nodules plus external model (ROI3) achieves an area under the ROC curve (AUC) of 0.93 (95% confidence interval 0.87 ~ 0.98) in the testing group, and have an accuracy of 0.859, sensitivity of 0.793, and specificity of 0.898 (Fig. 5A-C). Therefore, Nodules plus external model is the optimal prediction model. Then, the radiomic features of the optimal model were combined with clinical label (nodule type) to establish a tuberculosis prediction model, and the area under the ROC curve (AUC) of the validation group was 0.93 (95% confidence interval 0.88 ~ 0.98), accuracy was 0.870, sensitivity was 0.871 and specificity was 0.852, as shown in Fig. 6. Finally an radiomics nomogram which included radiomic features and clinical risk factor (nodule type) were established to predict the risk of pulmonary nodules being tuberculous (Fig. 7). Decision curve evaluation showed that the nomogram model had better clinical effectiveness at a risk threshold > 2% than the model without clinical label and radiomic feature model (Fig. 8).

4. Discussion

Recently,Most studies of pulmonary nodules focus on the discrimination of benign and malignant nodules, and pay insufficient attention to the classification and diagnosis of benign nodules[12, 13]. In order to reduce medical waste under the requirements of precision medical model, it is necessary to further classify and diagnose benign nodules. The most common types of benign pulmonary nodules are tuberculous nodules and non-tuberculosis inflammatory nodules[14]. The two types of pulmonary nodules are indistinguishable from the human eye on CT imaging[15]. Therefore, in this study, a model was established to distinguish the two types of benign nodules by means of imaging omics, and multiple layers of analysis were performed from the nodules themselves and the surrounding parenchyma. The

results showed that the AUC of the classification diagnosis model of nodule plus surrounding substance was 0.93, with high sensitivity and specificity, and the decision curve analysis had the best clinical efficacy, which could bring good help for clinical diagnosis and treatment plan.

By extracting high-throughput features of pulmonary nodules from CT images, radiomics quantifies deep-level feature studies to achieve the purpose of improving the accuracy of classification and diagnosis[16]. In recent study, researchers established a classification and diagnosis model for tuberculous nodules and lung adenocarcinoma using radiomics nomogram with better accuracy[17], the AUC of the training set, internal validation set and external validation set were 0.889, 0.879 and 0.809, respectively. This shows that radiomic has obvious advantages in the classification and diagnosis of different pathological types of pulmonary nodules. In the general data of pulmonary nodules in the two groups included in this study, although the difference of nodules type was statistically significant. Solid nodules were in the majority in the tuberculous nodules group, while solid and pure ground glass accounted for the same proportion in the non-tuberculous inflammatory nodules, there were still considerable difficulties in the classification and diagnosis of benign pulmonary nodules in practice. Therefore, the author believes that the image features can be excavated to further refine the classification of benign nodules, so as to guide the accurate diagnosis and treatment of pulmonary nodules.

Tuberculous nodules are a special type of inflammatory nodules whose pathological basis is quite different from that of conventional inflammatory nodules[18]. Solitary tuberculous nodules are often misdiagnosed as lung cancer, inflammatory nodules and other lesions because they are relatively rare and lack common tuberculosis features[19]. Tuberculous nodules are essentially granulomatous lesions, consisting mainly of fibrous tissue and its surrounding caseous necrosis, surrounded by granulation tissue[20]. Due to the proliferation of granulation tissue and the binding of the bronchial wall, there is less marginal infiltration. Inflammatory nodules are the result of a series of pathophysiological changes such as increased vascular permeability, inflammatory cell infiltration and serous exudation, and granulation tissue hyperplasia caused by pathogenic bacteria. Therefore, the edges of nodules are usually blurred and peripheral vascular congestion is increased[21]. Inflammatory nodules are generally considered to be of uniform density with no or rare bronchial aeration. Smooth margins, shallow lobules or long thick burrs are more common in benign lesions[22]. Therefore, the two types of nodules are significantly different in their central pathologic structure and peripheral exudation or microvascular infiltration. In this study, based on this pathological basis, the radiomics model was established at three levels: the nodule itself, the parenchyma zone around the nodule, and the nodule plus external expansion, which also put forward higher requirements for nodule segmentation.

Image segmentation plays a key role in the study of Radiomics[23]. Manual segmentation, semi-automatic segmentation and automatic segmentation are commonly used methods. Recent study showed that semi-automatic segmentation was more reliable for the radiomics parameters extracted from isolated pulmonary nodules and could provide objective and stable information for the classification model[24]. Other studies explored the inclusion of features around pulmonary nodules into deep learning tools to

evaluate benign and malignant pulmonary nodules, and found that the prediction model incorporating the surrounding parenchymal tissue with the 1/4 diameter band of the nodules had the best efficacy[25]. In this study, semi-automatic segmentation and manual correction of the nodule boundary were used to determine the pulmonary nodule as ROI1, and the 3mm parenchyma around the nodule was regarded as ROI2 by the expansion function of AK platform, and finally, the nodule plus expansion was treated as ROI3. Through the above method to achieve accurate segmentation.

The wavelet conversion feature can focus the energy of the original image on a small number of wavelet coefficients, and the decomposed wavelet coefficients have high local correlation on the detail components in three directions providing a strong condition for feature extraction[26]. Many studies have shown that wavelet changes can effectively remove the striation noise in CT and MRI images, which has obvious advantages in the classification and predictive analysis of radiomics and has been widely used in the study of radiomics and texture analysis[27, 28]. In this study, the selected features are mostly based on the wavelet characteristics of frequency transform. In the most efficient ROI3 model, 10 of the 11 radiomic features are wavelet features and the other one is original shape feature, which cover nodule shape feature, skewness of histogram, and high order features (GLCM, GLDM and GLRLM). The texture features of GLCM and GLDM can better display the texture information of nodules than histogram[29], and GLRLM has a significant effect in characterizing the consistency of nodules or speckled textures[30]. The features excavated in this study can reflect the texture difference between the two groups of nodules from different angles and dimensions of the image. At the same time, it was found that the nodule plus external model had the highest diagnostic efficacy, which also confirmed that the pathological basis of infiltration of the two types of nodules themselves and surrounding parenchymal is different, reflected in the image, that is, invisible radiomic features can distinguish the two types of nodule well.

The multiple Logistic regression analysis showed good performance in classification modeling[31, 32]. Recent research used this method to establish the benign and malignant differentiation model of pulmonary nodules, and the AUC reached 0.89[33]. In this study, the ROC curves of the training cohort and the validation cohort were modeled by 11 radiomic features of the Nodules plus external (ROI3), and the area under the ROC curves (AUC) of the two groups were both 0.93. It is worth noting that the AUC of the two groups was still 0.93 after the combination of radiomic features and clinical risk factor (nodule types), but the decision curves analysis (DCA) showed that the model combined with radiomic features had good clinical efficacy at the onset of risk. In this analysis, although the types of nodules observed by the naked eye of the traditional imaging methods were statistically significant in the two groups of different types of pulmonary nodules, they did not play a decisive role in the establishment of the classification model. However, radiomic features obtained after dimensionality reduction are useful for distinguishing non-tuberculous inflammatory pulmonary nodules from tuberculous nodules[34]. Therefore, this study can classify benign pulmonary nodules, non-tuberculous inflammatory pulmonary nodules and tuberculous nodules by using the radiomics nomogram model, which provides reference for the judgment of pathological types of benign pulmonary nodules before clinical operation in the future, and has certain guiding value.

This study has the following limitations: First, the sample size of this study is not large enough, which may affect the performance of the radiomics model, and the future work should focus on large-scale and multi-center studies. Secondly, some laboratory data samples of cases included in this study are too small, and important tuberculosis laboratory tests are often missing. At the same time, it is also the purpose of this study, if the nodules can be identified as tuberculosis in the early stage, the relevant clinical laboratory examination can be suggested. Although radiomics is widely used in medicine, the stability and redundancy of features are of concern. In future studies, these limitations will be reduced to a greater extent by expanding the sample size, establishing a standard database, integrating more laboratory examination indicators, and adopting artificial intelligence to automatically segment nodules, so as to make the classification model more efficient and applicable.

5. Conclusion

In summary, on the basis of both practicality and accuracy, this study established three radiomics models, including the pulmonary nodule, surrounding parenchymal and nodules plus expansion, analyzed the AUC of the three models, and found that the nodules plus expansion model is the optimal radiomics model for the classification and diagnosis of benign pulmonary nodules. The results of ROC and DCA analysis showed that compared with traditional imaging methods, this model has better predictive and classification value, and is expected to provide a basis for accurate diagnosis and treatment of pulmonary nodules.

6. Abbreviations

LDCT Low-dose Computed Tomography

ICC Intraclass Correlation Coefficient

mRMR Maximum relevance and Minimum redundancy

LASSO Least absolute shrinkage and selection operator

AUC Area Under Curve

CAD Computer-aided diagnosis

pGGN pure ground glass nodules

mGGN mixed ground glass nodules

SN solid nodules

DICOM Digital Imaging and Communications in Medicine

PACS Picture Archiving and Communication Systems

IBSI Image Biomarker Standardization Initiative

GLCM Gray-Level Co-occurrence Matrix

GLSZM Gray-Level Size Zone Matrix

GLRLM Gray-Level Run Length Matrix

NGTIM Neigbouring GrayTone Difference Matrix

GLDM Gray Level Dependence Matrix

DCA Decision curves analysis

7. Declarations

Acknowledgements

We thank the patients, the research team at GE Health, and the team of thoracic surgeons and pathologists who helped us with our research.

Authors' contributions

DZ and WJF participated in the conception, the design, coordination of thestudy, and manuscript preparing. ZS, YY and LW conceived of the study, andparticipated in its design. MK and LY participated in data collecting and ROI segmentation. WJH and HL participated in data analysis and model establishment and optimization. All authors readand approved the final manuscript.

Funding

This work was supported by the Clinical major innovative characteristic technology project of Second Affiliated Hospital of Army Medical University (2018JSLC0016) and the Medical&health care project of a Chinese institution (20BJZ05).The funders had no role in the study design, data collection and analysis, decision to publish, or preparationof the manuscript.

Availability of data and materials

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

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Second Affiliated Hospital of Army Medical University (2020-147-01). All subjects were exempted from informed consent.

Consent for publication

All participants in this study are exempt from informed consent according to the rules of the ethics committee.

Competing interests

Author Huan Liu was employed by company GE Healthcare. All other authors declare no competing interests.

8. References

1. Yankelevitz, D.F., et al., *CT Screening for Lung Cancer: Nonsolid Nodules in Baseline and Annual Repeat Rounds.* Radiology, 2015: p. 555-64.DOI: 10.1148/radiol.2015142554.
2. Baldwin, D.R., et al., *External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules.* Thorax, 2020. **75**(4): p. thoraxjnl-2019-214104.DOI: 10.1136/thoraxjnl-2019-214104.
3. Wilson, R. and A. Devaraj, *Radiomics of pulmonary nodules and lung cancer.* Translational Lung Cancer Research, 2017. **6**(1): p. 86.DOI: 10.21037/tlcr.2017.01.04.
4. Heo, J.N., et al., *Pulmonary tuberculosis: another disease showing clusters of small nodules.* ajr american journal of roentgenology, 2005. **184**(2): p. 639.DOI: 10.2214/ajr.184.2.01840639.
5. Lambin, P., et al., *Radiomics: Extracting more information from medical images using advanced feature analysis.* European Journal of Cancer, 2012. **43**(4): p. 441-446.DOI: 10.1016/j.ejca.2011.11.036.
6. Radiology, C.S.o., et al., *Expert consensus on the rule and quality control of pulmonary nodule annotation based on thoracic CT.* Chinese Journal of Radiology, 2019. **53**(1): p. 9-15.DOI: 10.3760/cma.j.issn.1005-1201.2019.01.004.
7. Hu, X., et al., *Computer-aided diagnosis of ground glass pulmonary nodule by fusing deep learning and radiomics features.*Phys Med Biol, 2021. **66**(6): p. 065015.DOI: 10.1088/1361-6560/abe735.
8. Zwanenburg, A., et al., *Image biomarker standardisation initiative.*Radiotherapy & Oncology, 2016.DOI: 10.1016/S0167-8140(18)31291-X.
9. Lee, D., S. Choi, and H.J. Kim, *High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains.* Medical Physics, 2018.DOI: 10.1002/mp.13258.
10. Wu, G., et al., *Preoperative CT-based radiomics combined with intraoperative frozen section is predictive of invasive adenocarcinoma in pulmonary nodules: a multicenter study.* European Radiology, 2020. **30**(5).DOI: 10.1007/s00330-019-06597-8.
11. Sun, Y., et al., *Radiomics for lung adenocarcinoma manifesting as pure ground-glass nodules: invasive prediction.* European Radiology, 2020(386–392).DOI: 10.1007/s00330-020-06776-y.
12. Ohno, Y., et al., *Differentiation of Benign from Malignant Pulmonary Nodules by Using a Convolutional Neural Network to Determine Volume Change at Chest CT.* Radiology, 2020. **296**(2): p.

432-443.DOI: 10.1148/radiol.2020191740.

13. Tang, S., et al., *Classification of Benign and Malignant Pulmonary Nodules Based on the Multiresolution 3D DPSECN Model and Semisupervised Clustering.* IEEE Access, 2021. **9**: p. 43397-43410.DOI: 10.1109/access.2021.3060178.
14. Yuan, X., et al., *Differentiation of malignant and benign pulmonary nodules with first-pass dual-input perfusion CT.* European Radiology, 2013. **23**(9): p. 2469-2474.DOI: 10.1007/s00330-013-2842-x.
15. Xia, X., et al., *Comparison and Fusion of Deep Learning and Radiomics Features of Ground-Glass Nodules to Predict the Invasiveness Risk of Stage-I Lung Adenocarcinomas in CT Scan.* Frontiers in Oncology, 2020. **10**.DOI: 10.3389/fonc.2020.00418.
16. Mackin, D., et al., *Measuring Computed Tomography Scanner Variability of Radiomics Features.* Investigative Radiology, 2015. **50**(11): p. 757.DOI: 10.1097/RLI.0000000000000180.
17. Feng, B., et al., *Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas.* Eur Radiol, 2020. **30**(12): p. 6497-6507.DOI: 10.1007/s00330-020-07024-z.
18. Jain, D., et al., *Pathology of Pulmonary Tuberculosis and Non-tuberculous Mycobacterial Lung Disease: Facts, Misconceptions, and Practical Tips for Pathologists.* Seminars in Diagnostic Pathology, 2017: p. S0740257017300874.DOI: 10.1053/j.semdp.2017.06.003.
19. Chong, et al., *The Necessity of Anti-Tuberculosis Therapy after Resection of Pulmonary Tuberculous Nodules: A Single Center Retrospective Study.* Annals of thoracic and cardiovascular surgery : official journal of the Association of Thoracic and Cardiovascular Surgeons of Asia, 2019.DOI: 10.5761/atcs.oa.19-00199.
20. Morey-Matamalas, A., et al., *Diagnostic Test Comparison of Tuberculous Lesions Found in Lymph Nodes From Slaughtered Cattle.* Journal of Comparative Pathology, 2018. **158**: p. 137-.DOI: 10.1016/j.jcpa.2017.10.131.
21. Wang, M., et al., *Correlation study between dual source CT perfusion imaging and the microvascular composition of solitary pulmonary nodules.* Lung Cancer, 2019. **130**: p. 115-120.DOI: 10.1016/j.lungcan.2019.02.013.
22. Hu, H., et al., *Multi-slice computed tomography characteristics of solitary pulmonary ground-glass nodules: Differences between malignant and benign.* Thoracic Cancer, 2016. **7**(1): p. 80-87.DOI: 10.1111/1759-7714.12280.
23. Ji, E.P., et al., *Diffusion and perfusion MRI radiomics obtained from deep learning segmentation provides reproducible and comparable diagnostic model to human in post-treatment glioblastoma.* European Radiology, 2020.DOI: 10.1007/s00330-020-07414-3.
24. Wu, W., et al., *Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: a case-control study.* European Radiology, 2019. **29**(11).DOI: 10.1007/s00330-019-06213-9.
25. Uthoff, J., et al., *Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT.* Med Phys, 2019. **46**(7): p. 3207-

3216.DOI: 10.1002/mp.13592.

26. Kan, E., J. Min, and J.C. Ye, *WaveNet: a deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction.* Medical Physics, 2016. **44**(10): p. e360.DOI: 10.1002/mp.12344.
27. Jooae, et al., *Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses.* Radiology, 2019. **292**(2): p. 365-373.DOI: 10.1148/radiol.2019181960.
28. Liao, F., et al., *Evaluate the Malignancy of Pulmonary Nodules Using the 3D Deep Leaky Noisy-or Network.* IEEE Transactions on Neural Networks & Learning Systems, 2017.DOI: 10.1109/TNNLS.2019.2892409.
29. Weng, Q., et al., *A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part-solid nodules.* Clinical Radiology, 2019. **74**(12).DOI: 10.1016/j.crad.2019.07.026.
30. Bogowicz, M., et al., *Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma.* Acta Oncologica, 2017. **56**(11): p. 1.DOI: 10.1080/0284186X.2017.1346382.
31. Jacobs, C., et al., *Solid, part-solid, or non-solid?: classification of pulmonary nodules in low-dose chest computed tomography by a computer-aided diagnosis system.* Investigative Radiology, 2015. **50**(3): p. 168-73.DOI: 10.1097/RLI.0000000000000121.
32. Xu, Y., et al., *Application of Radiomics in Predicting the Malignancy of Pulmonary Nodules in Different Sizes.* AJR Am J Roentgenol, 2019. **213**(6): p. 1213-1220.DOI: 10.2214/AJR.19.21490.
33. Garau, N., et al., *External validation of radiomics-based predictive models in low-dose CT screening for early lung cancer diagnosis.* Med Phys, 2020. **47**(9): p. 4125-4136.DOI: 10.1002/mp.14308.
34. Ather, S., T. Kadir, and F. Gleeson, *Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications.* Clin Radiol, 2020. **75**(1): p. 13-19.DOI: 10.1016/j.crad.2019.04.017.

Figures

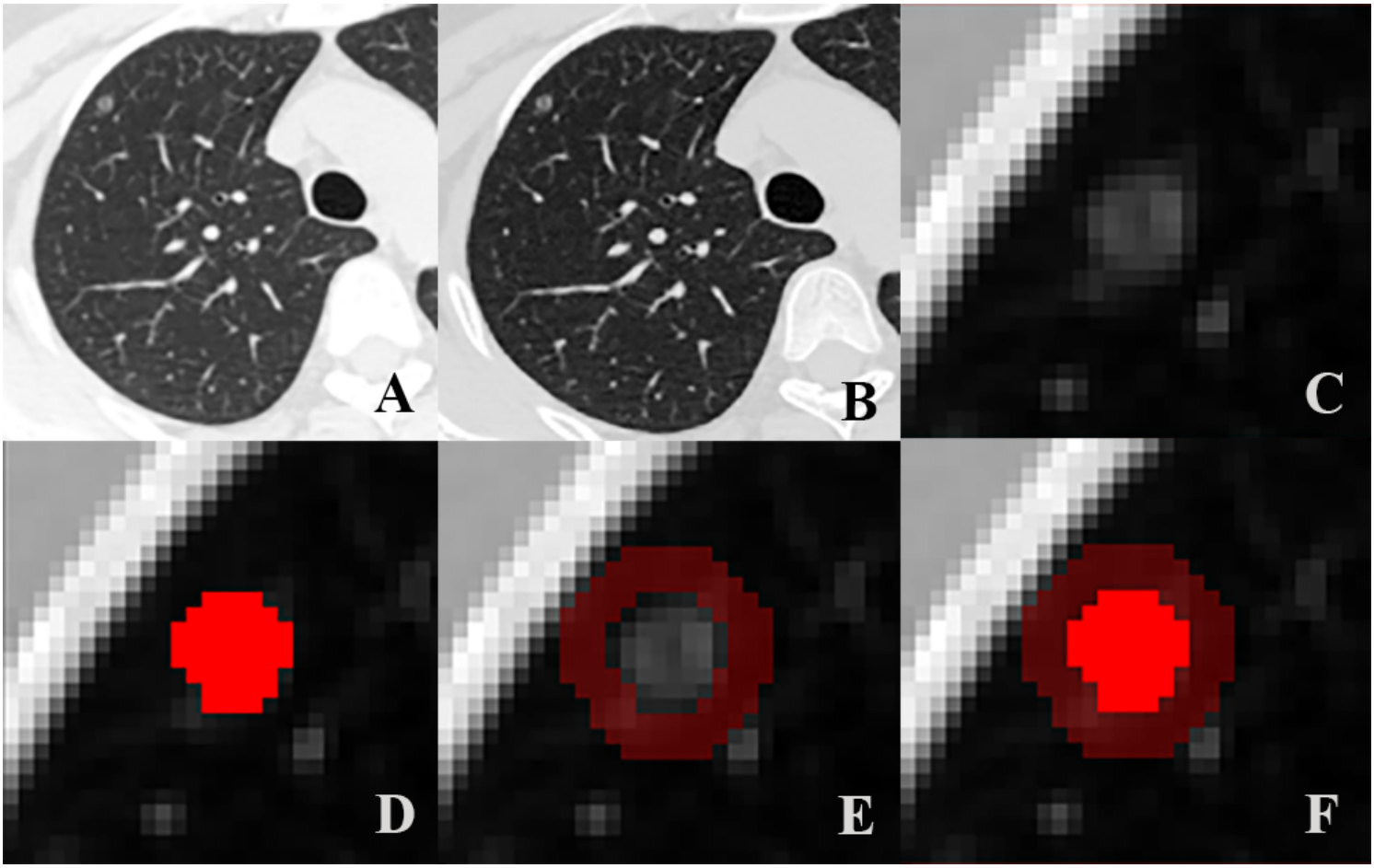

Figure 1

A-FSegmentation of pulmonary nodules. Figure A. is the original CT image of the lung window.Figure B. shows the resampled imagewith1 mm×1 mm × 1 mm. FigureC.Pulmonary nodules as the center, including the clip of surrounding normal lung tissue.Figure D. A seed point is placed at the center of the nodule and the rough boundary of the nodule is obtained by growth.Then, 2D tools were used to correct the cut images layer by layer, and lung nodules (ROI1) were segmented along the edges of nodules.Interference with peripheral pulmonary components such as blood vessels, cords, or pleura should be avoided as far as possible.Figure E. 3mm parenchymal band around the nodule (ROI2) was obtained by using the ROI extension function of AK software. Figure F. Pulmonary nodules with expanded parenchymal bands (ROI3).

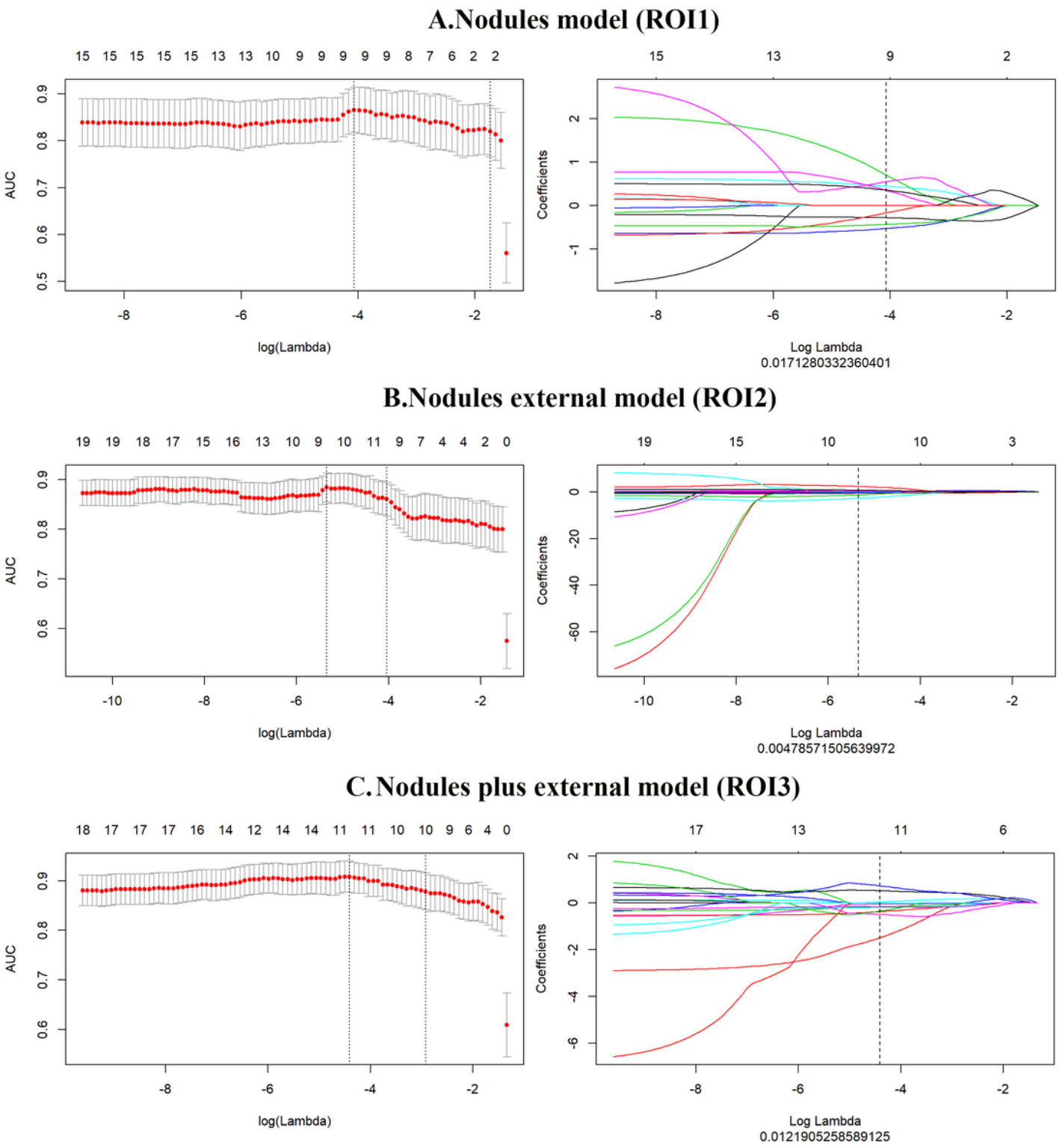

Figure 2

(A-C) The left figures, radiomic features were selected with the lowest binomial deviance. Coefficient λ was selected in the LASSO regression model using 10-fold cross-validation. According to the minimum criteria and the 1 standard error of the minimum criteria, dotted vertical lines were drawn at the optimal value. The right figures, LASSO coefficient profiles of radiomic features. Dotted vertical line was drawn at the optimal λ selected using 10-fold cross- validation.

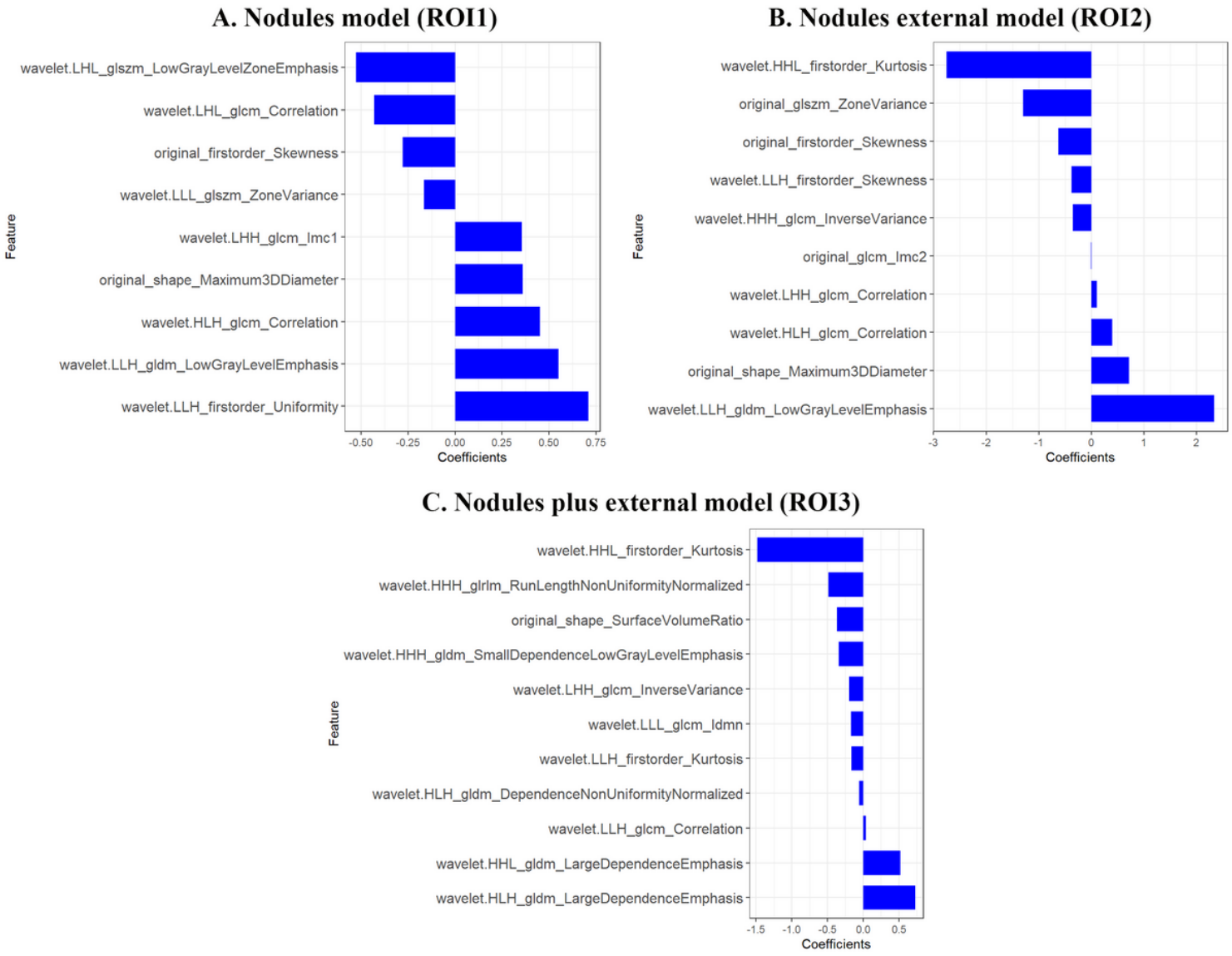

Figure 3

(A-C) The radiomic coefficients of each feature in the most predictive feature subset in the three models.

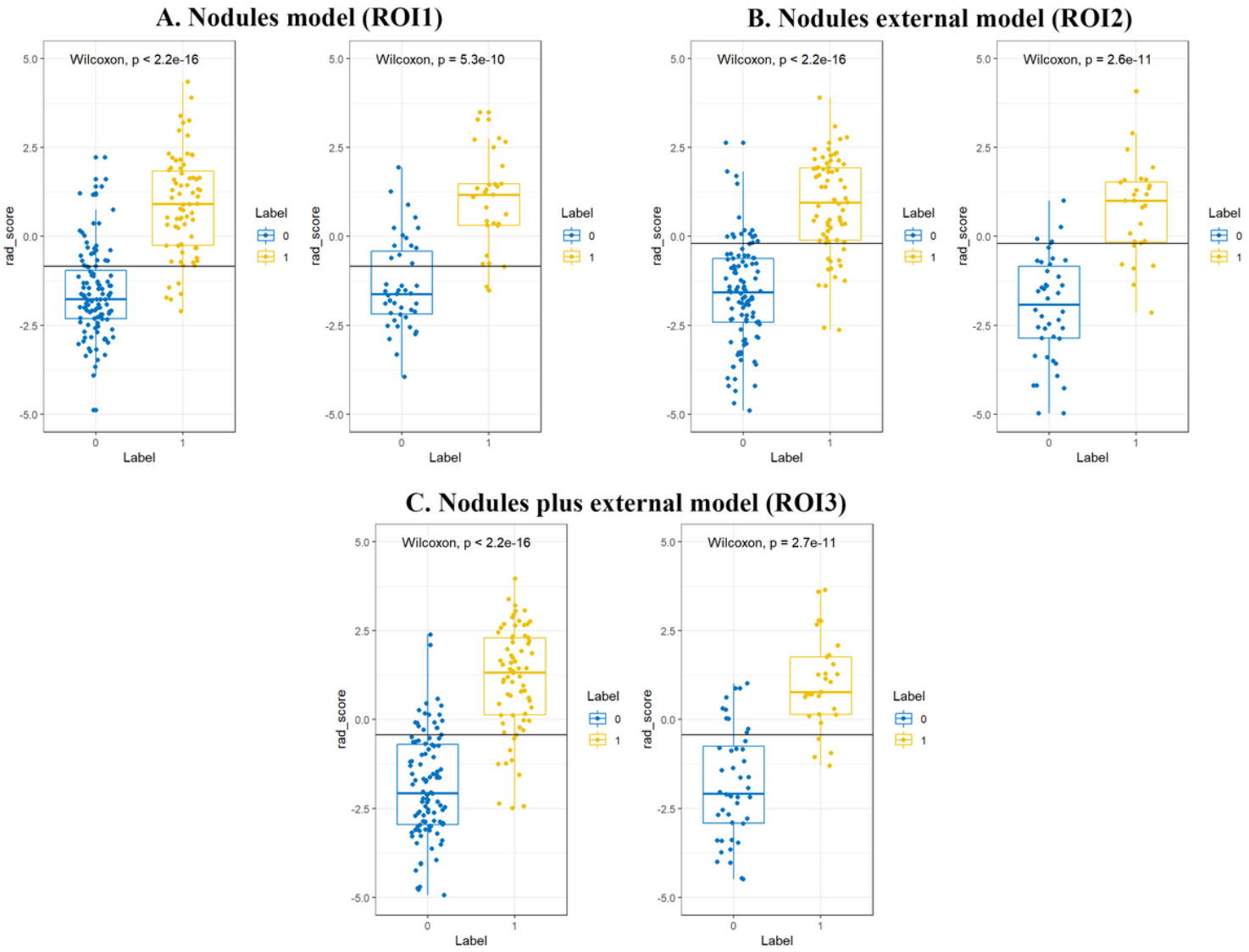

Figure 4

(A-C)Compared the RadScores from class 0 and class 1 on training group and test group in each models respectively.

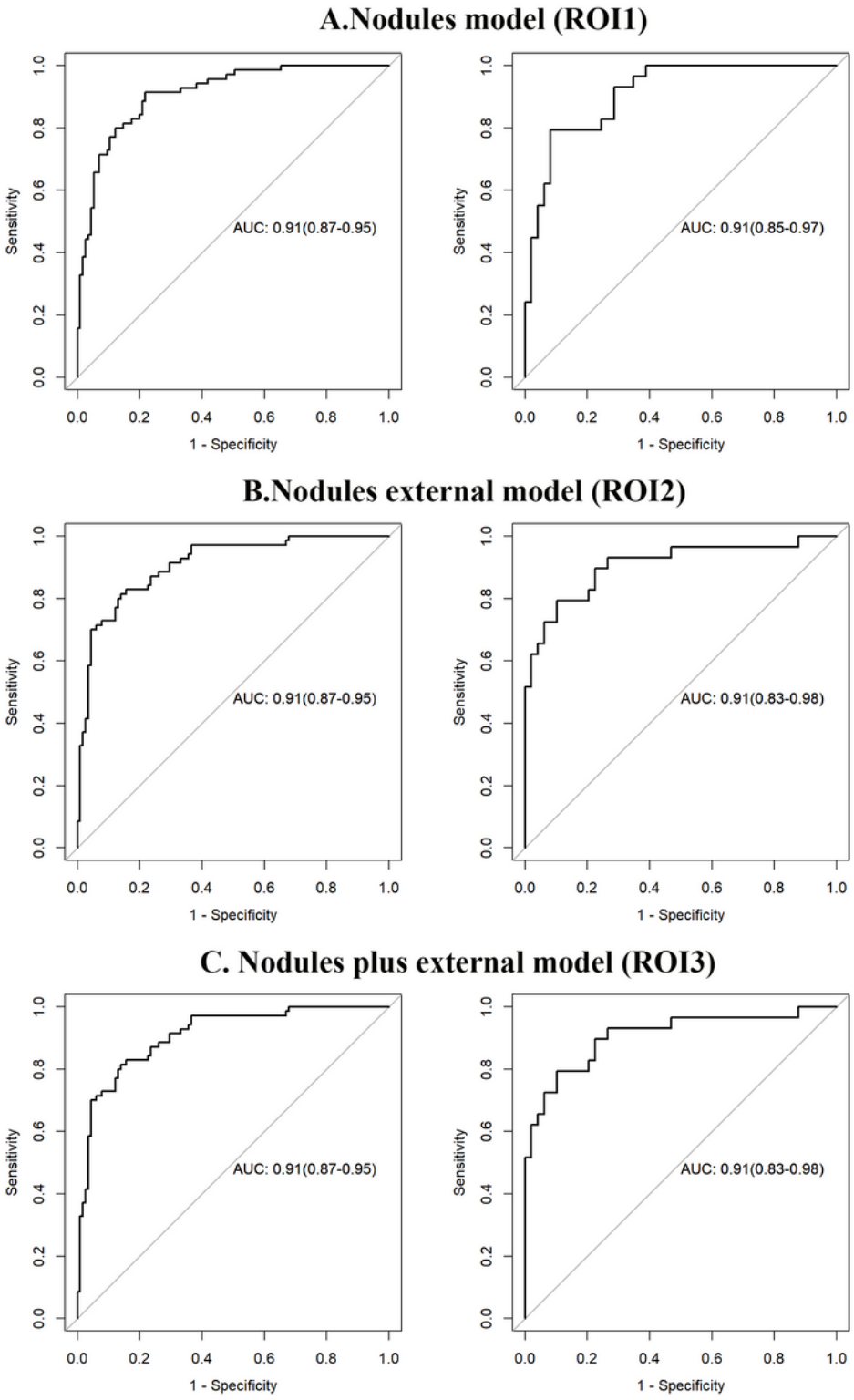

Figure 5

(A-C) ROC curve of training cohort and testing cohort in the three models.

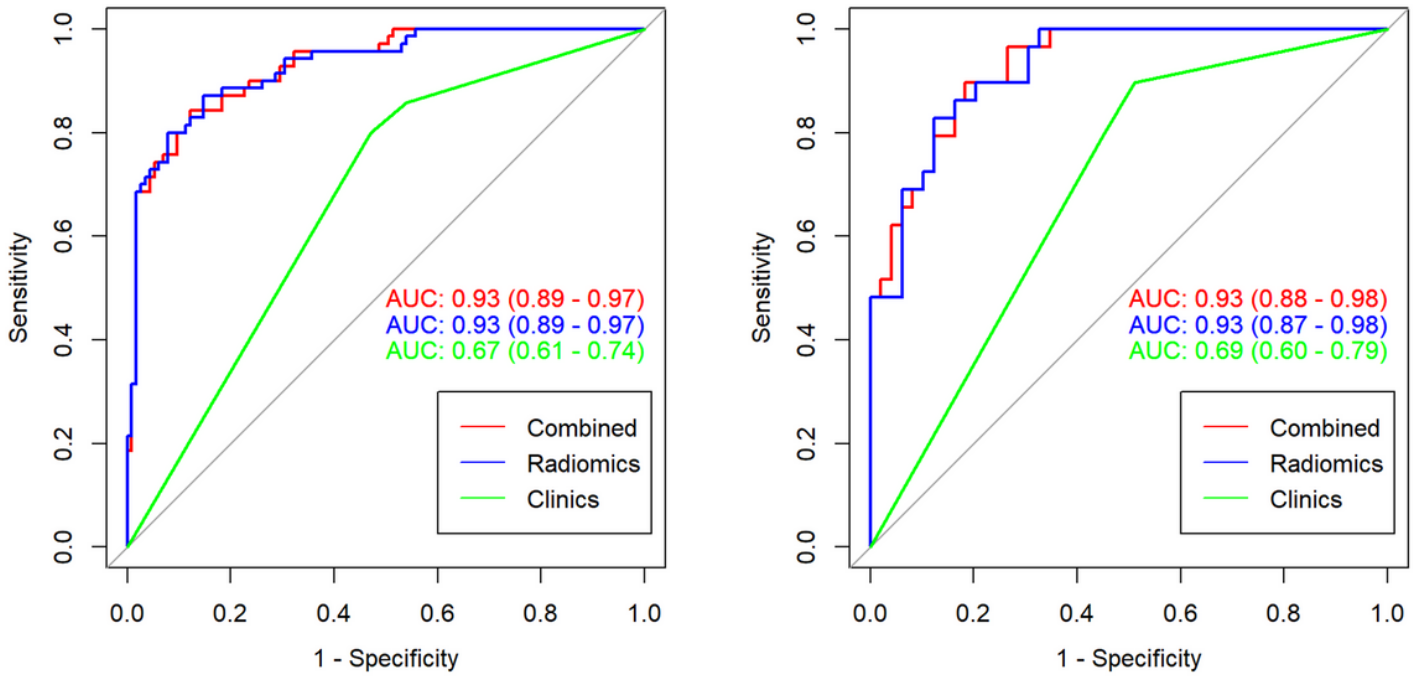

Figure 6

The ROC curves of the training set and the validation set for predicting the risk of tuberculosis nodules. The red curve represents the Nodules plus external model of radiomics labels combined with clinical risk factors (nodular type), with an AUC of 0.93. The blue curve represents the AUC of 0.93 when the radiomics label prediction model is used alone. The green curve represents the AUC of 0.67 and 0.69 when the clinical data prediction model was used alone.

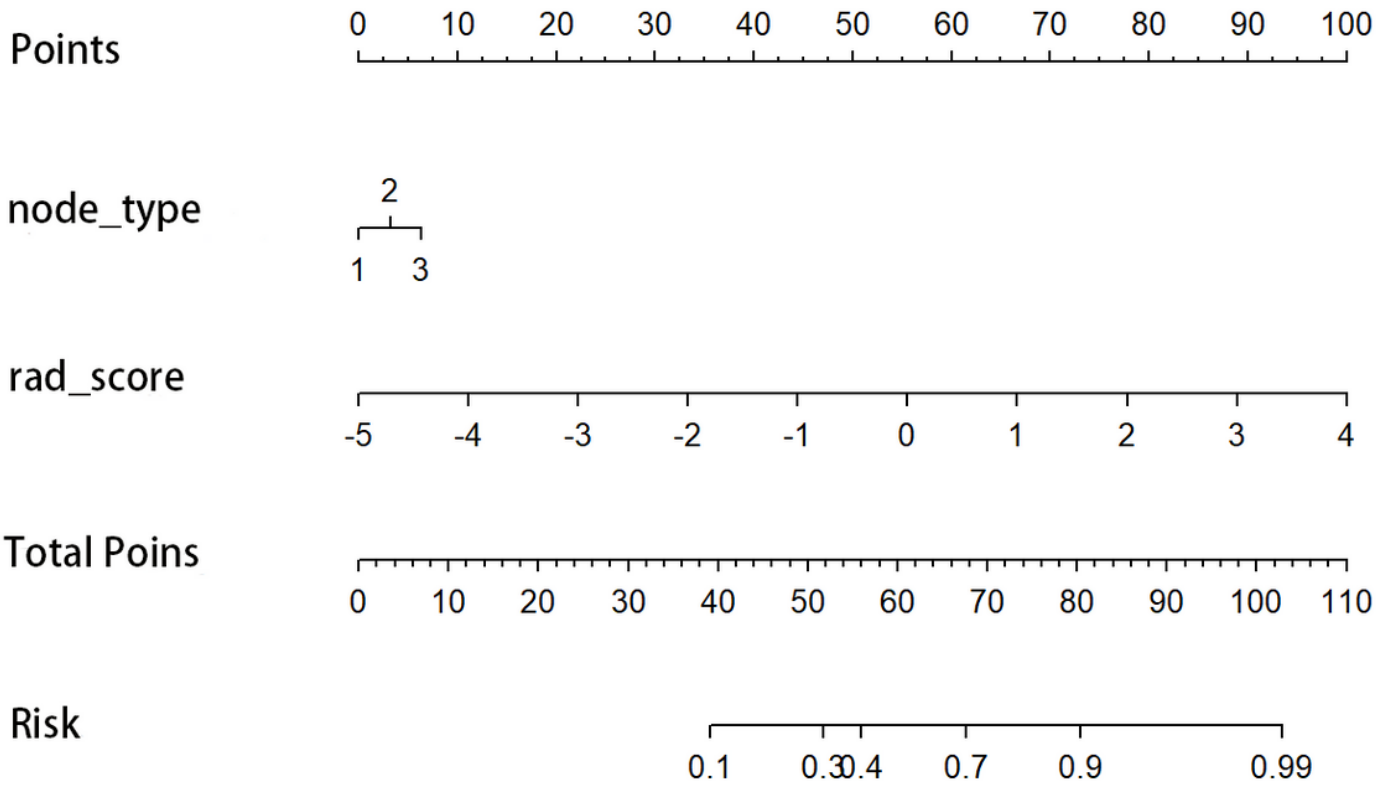

Figure 7

Radiomics nomogram. Risk factors included pulmonary nodule type and Radscore, where nodule type 1=pGGN, 2=mGGN, and 3=SN. Positioning was performed on the horizontal axis of Radscore and nodule type, and vertical lines were drawn to obtain the corresponding score values. The score values were added to the horizontal axis of the total score values for positioning, and vertical lines were drawn to the risk series number axis. The risk coefficient was used to predict the risk degree of tuberculosis nodules.

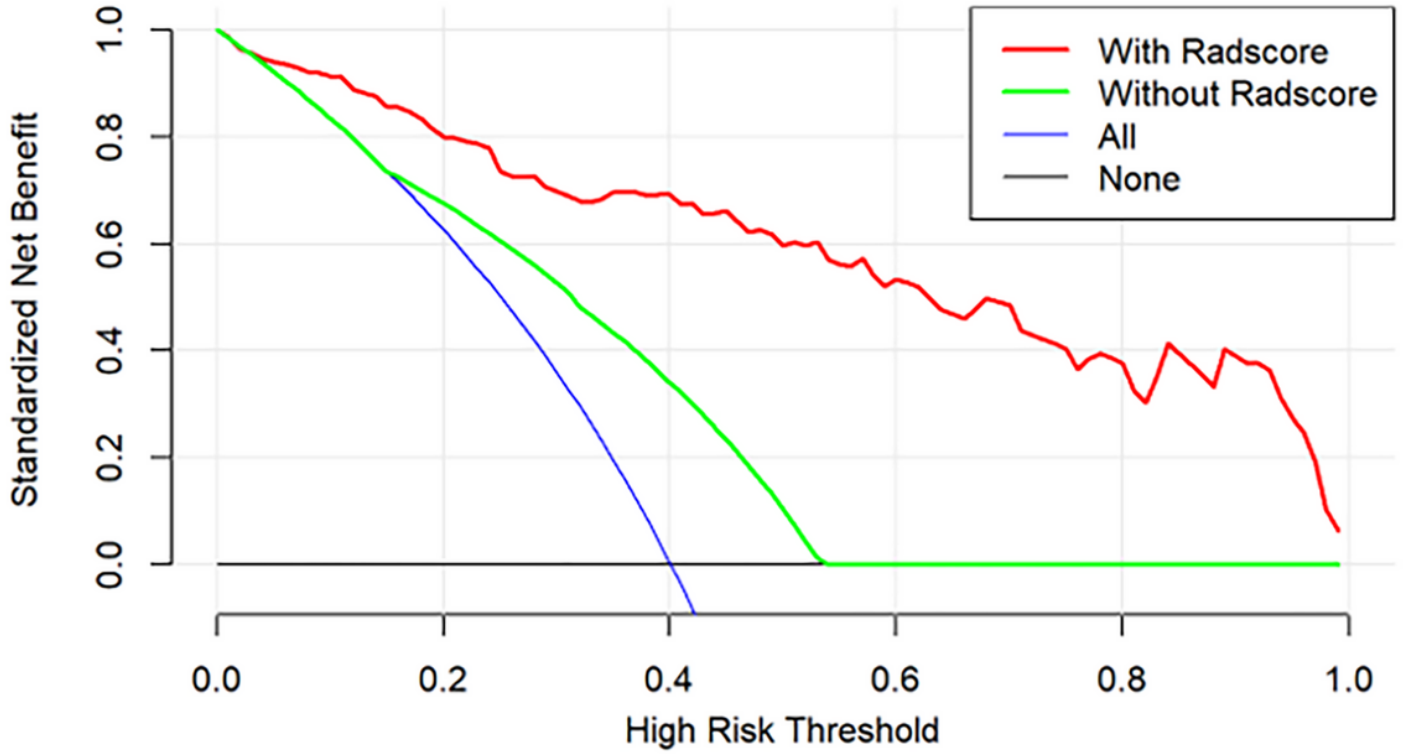

Figure 8

Decision curves analysis of the predictive model in all patients with pulmonary nodules. The decision curve represents the efficacy values under different risk thresholds. When the risk threshold is greater than 2%, the method of predicting the risk of tuberculosis nodules using the nomogram model is superior to identifying all nodules as tuberculosis nodules or non-tuberculosis inflammatory nodules, and also superior to the prediction method without radiomics label.