

## ORIGINAL ARTICLE

# Preoperative diagnosis of malignant pulmonary nodules in lung cancer screening with a radiomics nomogram

Ailing Liu<sup>1</sup> | Zhiheng Wang<sup>2</sup> | Yachao Yang<sup>3</sup> | Jingtao Wang<sup>2</sup> | Xiaoyu Dai<sup>2</sup> | Lijie Wang<sup>2</sup> | Yuan Lu<sup>2</sup> | Fuzhong Xue<sup>2,4</sup> 

<sup>1</sup>Department of Pulmonary and Critical Care Medicine, Weihai Municipal Hospital, Weihai, Shandong 264200, P. R. China

<sup>2</sup>Department of Biostatistics, School of Public Health, Shandong University, Jinan, Shandong 250002, P. R. China

<sup>3</sup>Department of Physical Examination, Weihai Municipal Hospital, Weihai, Shandong 264200, P. R. China

<sup>4</sup>Institute for Medical Dataology, Shandong University, Jinan, Shandong 250002, P. R. China

## Correspondence

Fuzhong Xue, Institute for Medical Dataology and Department of Biostatistics, School of Public Health, Shandong University, 12550 Erhuandong Road, Jinan 250002, Shandong, P. R. China.

Email: xuefzh@sdu.edu.cn

## Funding information

Key R & D project of Shandong Province, Grant/Award Number: 2018GSF118152

## Abstract

**Background:** Lung cancer is the most commonly diagnosed cancer worldwide. Its survival rate can be significantly improved by early screening. Biomarkers based on radiomics features have been found to provide important physiological information on tumors and considered as having the potential to be used in the early screening of lung cancer. In this study, we aim to establish a radiomics model and develop a tool to improve the discrimination between benign and malignant pulmonary nodules.

**Methods:** A retrospective study was conducted on 875 patients with benign or malignant pulmonary nodules who underwent computed tomography (CT) examinations between June 2013 and June 2018. We assigned 612 patients to a training cohort and 263 patients to a validation cohort. Radiomics features were extracted from the CT images of each patient. Least absolute shrinkage and selection operator (LASSO) was used for radiomics feature selection and radiomics score calculation. Multivariate logistic regression analysis was used to develop a classification model and radiomics nomogram. Radiomics score and clinical variables were used to distinguish benign and malignant pulmonary nodules in logistic model. The performance of the radiomics nomogram was evaluated by the area under the curve (AUC), calibration curve and Hosmer-Lemeshow test in both the training and validation cohorts.

**Results:** A radiomics score was built and consisted of 20 features selected by LASSO from 1288 radiomics features in the training cohort. The multivariate logistic model and radiomics nomogram were constructed using the radiomics score and patients' age. Good discrimination of benign and malignant pulmonary nodules was obtained from the training cohort (AUC, 0.836; 95% confidence interval [CI]: 0.793-0.879) and

**Abbreviations:** AUC, Area Under Curve; CI, confidence interval; CT, computed tomography; LASSO, least absolute shrinkage and selection operator; OR, odd ratio; ROC, receiver-operator characteristic; ROI, region of interest.

Ailing Liu and Zhiheng Wang contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Cancer Communications* published by John Wiley & Sons Australia, Ltd. on behalf of Sun Yat-sen University Cancer Center

validation cohort (AUC, 0.809; 95% CI: 0.745-0.872). The Hosmer-Lemeshow test also showed good performance for the logistic regression model in the training cohort ( $P = 0.765$ ) and validation cohort ( $P = 0.064$ ). Good alignment with the calibration curve indicated the good performance of the nomogram.

**Conclusions:** The established radiomics nomogram is a noninvasive preoperative prediction tool for malignant pulmonary nodule diagnosis. Validation revealed that this nomogram exhibited excellent discrimination and calibration capacities, suggesting its clinical utility in the early screening of lung cancer.

#### KEYWORDS

computed tomography, early screening, lung cancer, nomogram, pulmonary nodule, radiomics

## 1 | BACKGROUND

Global statistics have shown that lung cancer is the most commonly diagnosed cancer and the leading cause of cancer death in both men and women [1, 2]. Due to non-timely diagnosis, most lung cancers are in an advanced stage when diagnosed leading to a poor 5-year overall survival rate of only 16.3% [3], whereas the 10-year survival rate of patients with stage I lung cancer who underwent surgery is 92% [4]. As such, the early screening of high-risk patients and treatment of lung cancer can improve the patients prognoses and reduce the risk of lung cancer death [5-8].

Every year, a large number of people undergo health examinations in China. The pulmonary nodules high-risk population was screened by computed tomography (CT). With the wide application of high-resolution CT in recent years, small pulmonary nodules have been detected in as many as 14%-58% of screening cases [9]. Obtaining high-quality images with a minimum radiation dose is the future development direction in pulmonary nodule screening. Once pulmonary nodules are detected by CT examination, doctors can judge whether the nodules are benign or malignant based on their size, shape, boundary, density, relationship with blood vessels. According to the judgments of doctors, different treatment plans are adopted. In some cases, it is hard for young doctors who lack extensive experience to accurately determine whether the nodules are benign or malignant.

Radiomics is an emerging and promising method that can quickly extract large numbers of features from tomographic images using high-throughput calculations [10, 11]. Presently, radiomics can extract a large number of features from medical images, multiscale wavelet images, and Gaussian filtering images, and those features can usually be divided into first-order statistical features, shape-based features, and statistical-based texture features [10, 12]. Big data and an increasing number of pattern recognition tools have contributed to the development of radiomics.

Therefore, medical images are not only used as a visual tool but are considered as quantitative data that can be used for data mining [11, 13, 14]. The relationships between radiomics features and clinical features can be revealed by quantitative analysis [15]. Previous studies have shown that biomarkers based on radiomics features are associated with clinical outcomes. Radiomics features are used to develop diagnostic or prognostic models as a clinical tool for personalized diagnostics and clinical decision support [16, 17]. If the radiomics method is used in the early screening of lung cancer, it is possible to improve the diagnostic accuracy and the efficiency of clinical decision-making.

The present study aims to develop a radiomics model and a nomogram as clinical application tool for differentiating between benign and malignant pulmonary nodules in the early screening of lung cancer.

## 2 | MATERIALS AND METHODS

### 2.1 | Datasets

The clinical data and CT images of 875 patients who underwent physical examinations at the Weihai Municipal Hospital (Weihai, Shandong, China) between June 2013 and June 2018 were reviewed. Most patients were confirmed by surgical specimen, some patients were confirmed by preoperative biopsy, such as percutaneous puncture or bronchoscopic biopsy. Patients were eligible for inclusion if they met the following criteria: (1) physical examination by CT scan revealing pulmonary nodules; (2) asymptomatic at diagnosis; (3) aged over 18 years old; (4) no lung cancer or other malignant tumors in the past five years; (5) physical condition fit for biopsy or surgery; and (6) definite pathological diagnosis for benign or malignant nodules. The exclusion criteria were as follows: (1) nodules larger than 3 cm in diameter; (2) complicated with other pulmonary

lesions; (3) past history of lung disease; and (4) metastatic tumors.

Data related to age, sex, and pulmonary nodule location was collected. All benign and malignant pulmonary nodules patients were randomly divided into two sets (training and validation cohort) by simple random sampling. The training cohort was used to construct a model and the validation cohort was used to evaluate the model. The protocol of this study was approved by the Public Health Ethics Committee of Shandong University (Approval No. 20180801), and informed consent was waived.

## 2.2 | Image acquisition and radiomics feature extraction

The radiomics workflow is shown in Supplementary Figure S1. All patients underwent CT examination before percutaneous puncture or bronchoscopic biopsy. The CT machine used was SIEMENS SOMATOM Definition Flash system (Siemens Healthineers, Erlangen, Germany), and the CT scan parameters were as follows: 120 kV, 150 effective mAs, beam collimation of  $128 \times 0.6$  mm, pitch of 1.2, gantry rotation time of 0.5 s, and slice thickness of 1.0 mm for the reconstructed image. The CT images of each patient included the lungs window and mediastinum window. The lungs window was used to extract radiomics features.

The region of interest (ROI) of each CT image was segmented by one doctor (Dr. Ailing Liu, Department of Respiratory Internal Medicine, Weihai Municipal Hospital, Weihai, Shandong, China) using the published 3D Slicer software version 4.8.0 (National Institutes of Health, Bethesda, MD, USA) [18]. The CT images of each patient were semi-automatically segmented in the visible area of the lesion. The radiomics features were extracted using the PyRadiomics version 2.1 module of Python version 3.6 (Python Software Foundation, Beaverton, OR, USA) [19]. Radiomics features were subdivided into the following categories: first-order statistical features, shape-based features, and statistical-based texture features. In addition to the shape features, all features were computed on the original image or on a Gaussian- or Wavelet-filtered image. All of the features were defined as described by the Imaging Biomarkers Standardization Initiative [20].

## 2.3 | Statistical analysis

### 2.3.1 | Radiomics feature selection and radiomics score construction

The least absolute shrinkage and selection operator (LASSO) method was used to select the nonzero coefficient features

in the training cohort. The LASSO algorithm is a regression analysis method for feature selection and regularization that aims to enhance the prediction accuracy and interpretability of statistical models [21]. The formula to calculate the radiomics score can be found in the Supplementary Method.

### 2.3.2 | Predictive verification of the radiomics score

The established radiomics score and the clinical data, including age, sex, and pulmonary nodule location was used to construct a univariate logistic regression model and a multivariate logistic regression model with the training cohort. The final multivariate logistic regression was constructed using backward stepwise regression to obtain similar results with fewer variables. The prediction performance of the logistic regression model was quantified by receiver operating characteristic (ROC) curves and the area under the curve (AUC) in the training cohort and validation cohort. Statistical difference between ROC curves of training and validation cohort was tested using the DeLong method [22].

### 2.3.3 | Construction and evaluation of a radiomics nomogram

A nomogram was drawn based on the multivariate logistic regression model. The calibration curve and Hosmer-Lemeshow test were used to evaluate the validity of the radiomics nomogram.

Statistical analysis was performed using R software version 3.6 (The Free Software Foundation, Boston, MA, USA). The “glmnet” package was used to build the LASSO model. The “rms” package was used to draw the nomogram. The “pROC” package was used to draw the ROC curves. The “ggplot2” package was used to draw a waterfall plot and calibration curve. The “generalhoslem” package was used to perform the Hosmer-Lemeshow test. A two-sided  $P < 0.05$  was considered significant.

## 3 | RESULTS

### 3.1 | Patient characteristics

A total of 875 patients were enrolled. Six hundred and twelve (70%) patients were assigned to the training cohort, and 263 (30%) were assigned to the validation cohort. The clinical characteristics of the patients in the training and validation cohorts are shown in Table 1. Four hundred and ninety-nine (81.5%) patients in the training cohort and 207 (78.7%) in the validation cohort had malignant pulmonary nodules. There

**TABLE 1** Clinical characteristics of the 875 patients with pulmonary nodules in the training cohort and validation cohort

Characteristic	Training cohort ( <i>n</i> = 612)			Validation cohort ( <i>n</i> = 263)		
	Benign <i>n</i> (%)	Malignant <i>n</i> (%)	<i>P</i> value	Benign <i>n</i> (%)	Malignant <i>n</i> (%)	<i>P</i> value
Age (years)			< 0.001			0.546
	< 55	59 (52.2)	137 (27.5)	22 (39.3)	70 (33.8)	
	≥ 55	54 (47.8)	362 (72.5)	34 (60.7)	137 (66.2)	
Gender			0.113			0.537
	Female	60 (53.1)	308 (61.7)	32 (57.1)	130 (62.8)	
	Male	53 (46.9)	191 (38.3)	24 (42.9)	77 (37.2)	
Location of pulmonary nodule			0.030			0.506
In lung	Right upper	32 (28.3)	190 (38.1)	12 (21.4)	65 (31.4)	
	Right middle	10 (8.8)	37 (7.4)	5 (8.9)	22 (10.6)	
	Right lower	30 (26.5)	85 (17.0)	10 (17.9)	35 (16.9)	
	Left upper	18 (15.9)	112 (22.4)	15 (26.8)	50 (24.2)	
	Left lower	23 (20.4)	75 (15.0)	14 (25.0)	35 (16.9)	

Note: *P* values were obtained from the univariate association analyses between the pulmonary nodules and each clinical characteristic.

was no significant difference in the rate of patients with malignant tumors between the two cohorts ( $P = 0.380$ ,  $\chi^2$  test).

### 3.2 | Establishment and validation of the radiomics score

A total of 1288 radiomics features were extracted from each patient's lung CT images. Detailed information about the radiomics features can be found in Supplementary Method and Supplementary Table S1.

The LASSO logistic regression model identified a total of 20 radiomics features with nonzero coefficients in the training cohort (Figure 1A and B). The 20 radiomics features are summarized in Supplementary Table S2. The radiomics scores of patients with malignant pulmonary nodules were usually higher than those with benign pulmonary nodules. In the training cohort, the radiomics scores (median [interquartile range]) of the malignant patient group (1.906 [1.379-2.416]) were significantly different ( $P < 0.001$ ) from those of the benign patient group (0.826 [0.338-1.271]). Additionally, in the validation cohort, the radiomics scores of the malignant patient group (2.035 [1.500-2.496]) were also significantly different ( $P < 0.001$ ) from those of the benign patient group (0.886 [0.451-1.575]).

As shown in Figure 1C and D, most patients with malignant pulmonary nodules (88.4%, 183/207) were diagnosed by the cutoff line of the radiomics score in the validation cohort. Based on the maximum Youden index in the training cohort, we found that the cutoff line of the radiomics score was 0.981.

To further verify the validity of the radiomics score, we compared the ROC curve of the logistic regression model with the clinical variables (age, sex, and pulmonary nodule location) to that of the hybrid logistic regression model. Figure 2

shows that the AUC of the hybrid logistic model was higher than that of the clinical variables.

### 3.3 | Validation and performance of the logistic model and radiomics nomogram

Table 2 shows the results of the univariate and multivariate logistic regression analyses. The backward logistic regressions which identified the radiomics score (odds ratio [OR], 5.77; 95% confidence interval [CI], 4.03-8.27;  $P < 0.001$ ) and age (OR, 1.04; 95% CI, 1.01-1.06;  $P = 0.002$ ) as independent predictors.

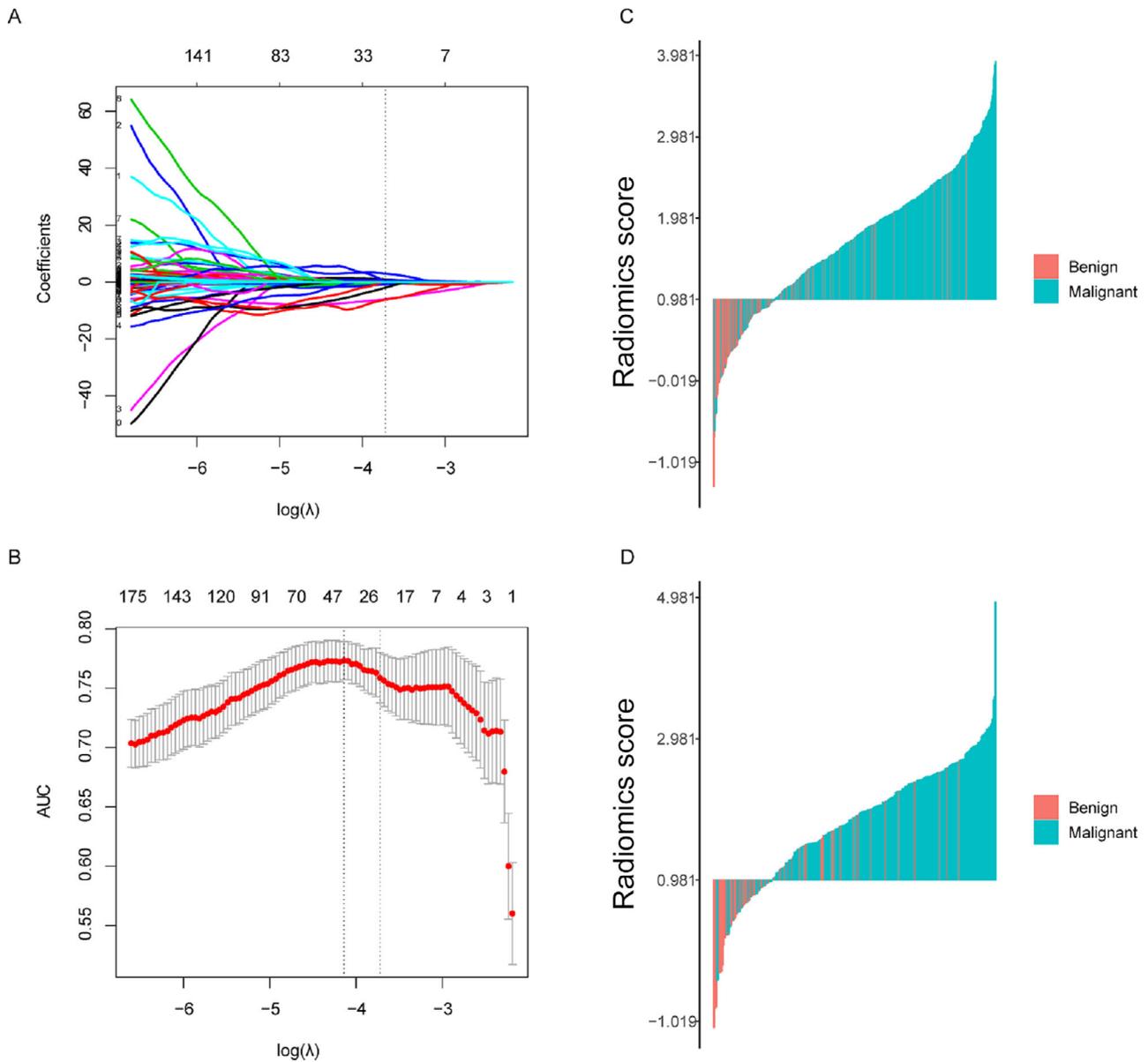
The AUC of the logistic regression model of the training cohort was 0.836 (Figure 3A, 95% CI, 0.793-0.879) and that in the validation cohort was 0.809 (Figure 3B, 95% CI, 0.745-0.872). As shown in Figure 3C, there was no significant difference ( $P = 0.485$ ) between these two ROC curves.

The radiomics nomogram was shown in Figure 4A. The calibration curve showed good calibration for the probability of malignant pulmonary nodules, the Hosmer-Lemeshow test showed a nonsignificant goodness of fit in the training cohort ( $P = 0.765$ , Figure 4B) and in the validation cohort ( $P = 0.064$ , Figure 4C).

## 4 | DISCUSSION

In this retrospective study, we established a logistic regression model and radiomics nomogram that could distinguish between benign and malignant pulmonary nodules during the early screening of lung cancer.

Although traditional medical diagnostic model can grasp the key information of identification between benign and



**FIGURE 1** Radiomics feature selection using the LASSO logistic regression model. A. The curve of the coefficient path of 1288 radiomics features in the training cohort. The dashed vertical line was drawn using the value of the selected  $\log(\lambda)$  in the 10-fold cross-validation in B, and there are 20 features with nonzero coefficients. B. The adjustment penalty parameter  $\lambda$  was selected in the LASSO model by 10-fold cross-validation based on the error within one standard error range of the minimum. The AUC values from the LASSO regression cross-validation process were plotted as a function of  $\log(\lambda)$ . The y-axis represents the AUC value. The lower x-axis represents  $\log(\lambda)$ . The number above the x-axis represents the average of the predictors. The red dot indicates the AUC value of each model with a given  $\lambda$ , and the vertical bar of the red dot indicates the upper and lower limits of the deviation. The vertical black line defines the best value for  $\lambda$ , where the model can provide the best result for the data. The best  $\lambda$  value was 0.024,  $\log(\lambda) = -3.730$ , and a total of 20 variables were selected. C-D. The waterfall plot of the training cohort (C) and validation cohort (D) was used to visualize the distribution of the radiomics score and the benign and malignant state of the pulmonary nodules of individual patients. The best cutoff value was 0.981. Abbreviations: LASSO, least absolute shrinkage and selection operator; AUC, area under the curve

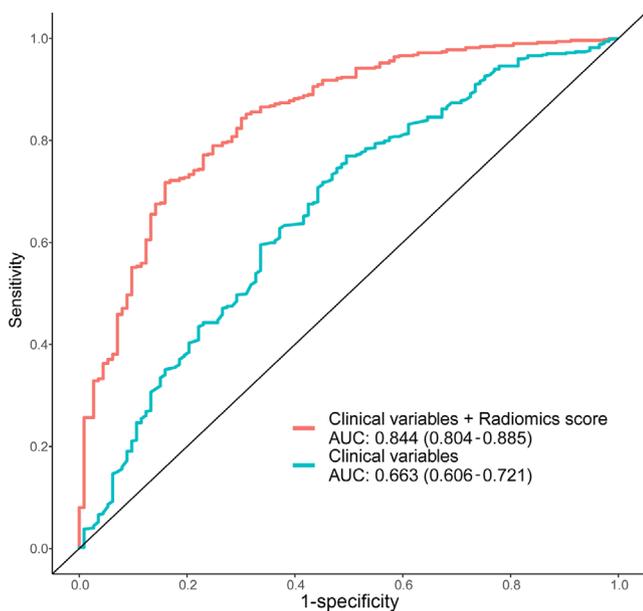
malignant pulmonary nodules, it inevitably has the disadvantages of bias in diagnosing the nodules. Additionally, in current imaging diagnostics, the analysis of medical images is performed by simple statistics of CT values and other values, but no in-depth and detailed analyses are performed. In fact, clinical data, CT and other imaging histological markers contain important information about pulmonary nodules.

Previously, a large number of studies have confirmed that radiomics is effective for the diagnosis and prognosis of lung cancer and pulmonary nodules [12, 23, 24]. Previous studies used statistical methods such as Kaplan-Meier survival curves and gene-set enrichment analysis to analyze the association of radiomics features with tumor phenotype, prognosis, and even gene expression [12]. However, you will find that these

**TABLE 2** Univariate and multivariate logistic regression analysis of the predicted factors for pulmonary nodules of patients in the training cohort

Characteristic		Univariate analysis		Multivariate analysis	
		OR (95% CI)	P value	OR (95% CI)	P value
Radiomics score		5.89 (4.12-8.41)	< 0.001	5.77 (4.03-8.27)	< 0.001
Age (years)	< 55	1.00		1.00	
	≥ 55	1.05 (1.02-1.07)	< 0.001	1.04 (1.01-1.06)	0.002
Gender	Male	1.00		–	–
	Female	0.70 (0.47-1.06)	0.092	–	–
Location of pulmonary nodule in lung	Right upper	1.00		–	–
	Right middle	0.62 (0.28-1.38)	0.242	–	–
	Right lower	0.48 (0.27-0.84)	0.010	0.58 (0.30-1.12)	0.105
	Left upper	1.05 (0.56-1.95)	0.883	–	–
	Left lower	0.55 (0.30-1.00)	0.050	–	–

Note: –, not available. The variables were eligible for inclusion in the multivariate analysis if  $P < 0.05$  in univariate analysis, so the OR and  $P$  values were not available. Abbreviations: OR, odd ratio; CI, confidence interval.



**FIGURE 2** Validation of the radiomics score. The comparison of the ROC curve of the logistic regression model with the clinical variables (red line) and the ROC curve of the logistic model with clinical variables along with the radiomics score (green line). Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve

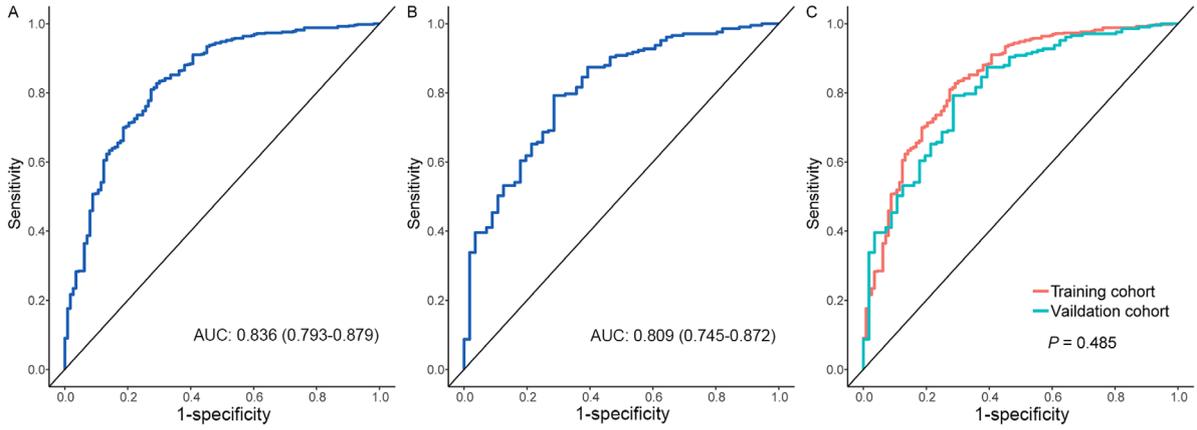
previous studies used fewer radiomics features; too few radiomics features will make a decline in effectiveness and stability. In this study, we used the LASSO model to select features from 1288 radiomics features, including first-order statistical features, shape-based features, statistical-based texture features and Gaussian and Wavelet information that can improve the stability of the radiomics model.

From the waterfall map in the validation cohort (Figure 1D), 88.4% of malignant pulmonary nodules could be correctly classified, showing that the radiomics

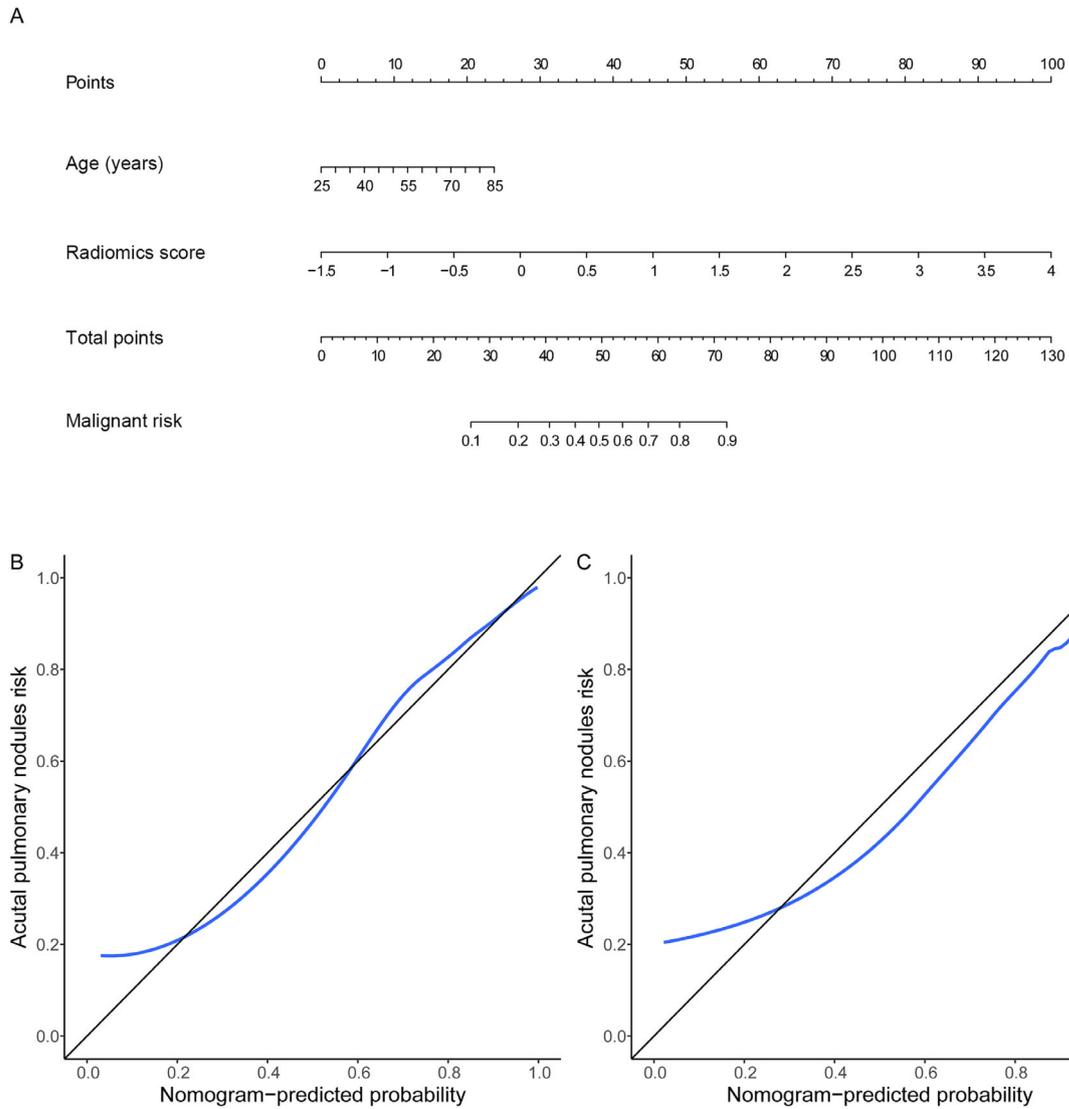
score was able to effectively distinguish benign and malignant pulmonary nodules in the early screening of lung cancer. By comparing the ROC curves of the radiomics score and traditional risk factors, the AUC increased from 0.663 to 0.844, indicating radiomics score have advantages over traditional risk factors.

Finally, a radiomics nomogram based on a logistic regression model was developed and can be used as a clinically easy-to-use tool. Doctors can use this radiomics nomogram to calculate the risk of malignant pulmonary nodules quickly and easily after further validation in multicenter cohorts and in different populations. In the validation cohort, the calibration curve demonstrated a good degree of coincidence with the 45° perfect prediction line, and the  $P$  value of the Hosmer-Lemeshow test statistic was nonsignificant, which indicates that the calibration of the radiomics nomogram based on the agreement between the predicted and observed results were reliable.

LASSO-logistic regression and nomogram have been applied to the prediction of lymph node metastasis in colorectal cancer and bladder cancer and in the survival prediction of non-small cell lung cancer [25-27]. The positive results of these articles reveal the effectiveness of radiomics in similar studies. However, the application of radiomics in early lung cancer screening mostly uses machine learning for pulmonary nodule localization or building a discriminant model [28, 29]. Since clinical application tools are not available, this research is difficult to widely spread to clinical applications. Additionally, almost no study has applied radiomics and nomograms in the diagnosis of pulmonary nodules. Considering the important role of early screening for lung cancer, there is great potential to apply radiomics and nomograms to improve the survival rate of lung cancer patients. Therefore, this article investigated the application of radiomics and developed a radiomics nomogram to assist doctors in diagnosing



**FIGURE 3** The ROC curve of the logistic regression model. A-B shows the ROC curves of the multivariate logistic regression model with radiomics score and age in the training (A) and validation (B) cohorts. C shows the test between the two ROC curves using the DeLong method



**FIGURE 4** Radiomics nomogram and calibration curves. A. Radiomics nomogram for predicting benign and malignant pulmonary nodules. B-C. Calibration curves of the radiomics nomogram in the training (B) and validation (C) cohorts. The calibration curve describes the calibration of the nomogram based on the agreement between the prediction of the benign and malignant pulmonary nodules and the observed actual benign and malignant results. The black line represents the perfect prediction, and the blue line represents the prediction performance of the radiomics nomogram

pulmonary nodules. As a noninvasive diagnostic tool, the nomogram will contribute to the guidance of surgical treatments and clinical decision-making.

In addition, there are some notable points and limitations in this article. Firstly, most benign patients don't choose to have surgery or biopsy, we can't obtain pathological diagnosis, so the proportion of malignant cases in this article is high. We chose the logistic model that is not sensitive to unbalanced data. Secondly, Referring to the 2017 Guidelines for the management of incidental pulmonary nodules [30], some traditional key risk factors, such as nodule size, morphology, and multiplicity, can be described by radiomics features. However, because our research focused on the relationship between CT and pathology, other traditional key risk factors such as family history, smoking and drinking history were missing when we collected data. In a future study, we hope to add these variables to improve the model. At the same time, the model performs well with good results in the absence of the traditional key variables of lung cancer, which also demonstrates the importance of radiomics. Finally, this study collected data from a single-center, although we divided the validation set to evaluation the stability of model, obviously data from multicenter cohorts and different populations are better, so multicenter research is the direction of further research.

## 5 | CONCLUSIONS

In this retrospective study, we constructed a model for distinguishing between benign and malignant lung nodules in early lung cancer screening. A nomogram was then built and proposed as a clinical-assisting diagnostic tool. It is noninvasive, inexpensive and can simplify decision-making efficiency to improve the lung cancer survival rate.

## ACKNOWLEDGEMENTS

We would like to thank all authors, reviewers and editors for their critical discussion of this manuscript, and apologize to those not mentioned due to space limitations.

## DECLARATIONS: ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The protocol of this study was approved by the Public Health Ethics Committee of Shandong University (Approval No. 20180801) and the informed consent was waived.

## CONSENT FOR PUBLICATION

Not applicable.

## AVAILABILITY OF DATA AND MATERIALS

The datasets are available from the corresponding author on reasonable request.

## COMPETING INTERESTS

The authors declare that they have no competing interests.

## FUNDING

This work was funded by Key R & D project of Shandong Province (2018GSF118152).

## AUTHORS' CONTRIBUTIONS

AL and ZW drafted this manuscript. AL, YY, LW, and YL collected the data. ZW, JW, XD analyzed the data. FX conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

## ORCID

Fuzhong Xue  <https://orcid.org/0000-0003-0378-7956>

## REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. <https://doi.org/10.3322/caac.21492>.
2. Feng R-M, Zong Y-N, Cao S-M, Xu R-H. Current cancer situation in China: good or bad news from the 2018 Global Cancer Statistics? *Cancer Commun*. 2019;39(1):22. <https://doi.org/10.1186/s40880-019-0368-6>.
3. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143(5):e1S–e29S.
4. Zheng R, Zeng H, Zhang S, Chen W. Estimates of cancer incidence and mortality in China, 2013. *Chin J Cancer*. 2017;36(1):66. <https://doi.org/10.1186/s40880-017-0234-3>.
5. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and Harms of CT Screening for Lung Cancer: A Systematic Review. *JAMA*. 2012;307(22):2418. <https://doi.org/10.1001/jama.2012.5521>.
6. Miser WF. Cancer screening in the primary care setting: the role of the primary care physician in screening for breast, cervical, colorectal, lung, ovarian, and prostate cancers. *Primary Care: Clinics in Office Practice*. 2007;34(1):137–67.
7. Wang Z, Wang Y, Huang Y, Xue F, Han W, Hu Y, et al. Challenges and research opportunities for lung cancer screening in China. *Cancer Commun*. 2018;38(1):34. s40880-018-0305-0. <https://doi.org/10.1186/s40880-018-0305-0>.
8. Wang Z, Han W, Zhang W, Xue F, Wang Y, Hu Y, et al. Mortality outcomes of low-dose computed tomography screening for lung cancer in urban China: a decision analysis and implications for practice. *Chin J Cancer*. 2017;36(1):57. <https://doi.org/10.1186/s40880-017-0221-8>.
9. Ellis MC, Hessman CJ, Weerasinghe R, Schipper PH, Vetto JT. Comparison of pulmonary nodule detection rates between preoperative CT imaging and intraoperative lung palpation. *Am J Surg*. 2011;201(5):619–22.
10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–6. <https://doi.org/10.1016/j.ejca.2011.11.036>.

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563–77. <https://doi.org/10.1148/radiol.2015151169>.
12. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5(1). <https://doi.org/10.1038/ncomms5006>.
13. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin*. 2019;caac.21552. <https://doi.org/10.3322/caac.21552>.
14. Aerts HJWL. Data Science in Radiology: A Path Forward. *Clin Cancer Res*. 2018;24(3):532–4. <https://doi.org/10.1158/1078-0432.CCR-17-2804>.
15. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res*. 2017;23(15):4259–69. <https://doi.org/10.1158/1078-0432.CCR-16-2910>.
16. Lee G, Lee HY, Park H, Schiebler ML, Beek EJR, Ohno Y, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol*. 2016;86:297.
17. Kotrotsou A, Zinn PO, Colen RR. Radiomics in Brain Tumors. *Magn Reson Imaging Clin N Am*. 2016;24(4):719–29. <https://doi.org/10.1016/j.mric.2016.06.006>.
18. Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. *Intraoperative imaging and image-guided therapy*. Springer; 2014;277–89.
19. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–e7.
20. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv:161207003 [cs]*. 2016.
21. Caron F. Regression shrinkage and selection via the Lasso. Robert Tibshirani. 1996;13.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988;44(3):837. <https://doi.org/10.2307/2531595>.
23. Kalpathy-Cramer J, Mamomov A, Zhao B, Lu L, Cherezov D, Napel S, et al. Radiomics of lung nodules: a multi-institutional study of robustness and agreement of quantitative imaging features. *Tomography*. 2016;2(4):430.
24. Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. *Transl Lung Cancer Res*. 2017;6(1):86.
25. Huang Y-q, Liang C-h, He L, Tian J, Liang C-s, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol*. 2016;34(18):2157–64. <https://doi.org/10.1200/JCO.2015.65.9128>.
26. Yang L, Yang J, Zhou X, Huang L, Zhao W, Wang T, et al. Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur Radiol*. 2019;29(5):2196–206. <https://doi.org/10.1007/s00330-018-5770-y>.
27. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. *Clin Cancer Res*. 2017;23(22):6904–11. <https://doi.org/10.1158/1078-0432.CCR-17-1510>.
28. Armato SG, Altman MB, Wilkie J, Sone S, Li F, Doi K, et al. Automated lung nodule classification following automated nodule detection on CT: A serial approach. *Med Phys*. 2003;30(6):1188–97. <https://doi.org/10.1118/1.1573210>.
29. Li Q, Balagurunathan Y, Liu Y, Qi J, Schabath MB, Ye Z, et al. Comparison Between Radiological Semantic Features and Lung-RADS in Predicting Malignancy of Screen-Detected Lung Nodules in the National Lung Screening Trial. *Clin Lung Cancer*. 2018;19(2):148–56.e3. <https://doi.org/10.1016/j.clcc.2017.10.002>.
30. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology*. 2017;284(1):228–43. <https://doi.org/10.1148/radiol.2017161659>.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Liu A, Wang Z, Yang Y, et al. Preoperative diagnosis of malignant pulmonary nodules in lung cancer screening with a radiomics nomogram. *Cancer Communications*. 2020;40:16–24. <https://doi.org/10.1002/cac2.12002>