BMC Medical Imaging

RESEARCH

Open Access

The scoring system combined with radiomics and imaging features in predicting the malignant potential of incidental indeterminate small (<20 mm) solid pulmonary nodules

Bai-Qiang Qu[1], Yun Wang[2], Yue-Peng Pan[3], Pei-Wei Cao[3] and Xue-Ying Deng[3*]

Abstract

Objective Develop a practical scoring system based on radiomics and imaging features, for predicting the malignant potential of incidental indeterminate small solid pulmonary nodules (IISSPNs) smaller than 20 mm.

Methods A total of 360 patients with malignant IISSPNs ($n=213$) and benign IISSPNs ($n=147$) confirmed after surgery were retrospectively analyzed. The whole cohort was randomly divided into training and validation groups at a ratio of 7:3. The least absolute shrinkage and selection operator (LASSO) algorithm was used to debase the dimensions of radiomics features. Multivariate logistic analysis was performed to establish models. The receiver operating characteristic (ROC) curve, area under the curve (AUC), 95% confidence interval (CI), sensitivity and specificity of each model were recorded. Scoring system based on odds ratio was developed.

Results Three radiomics features were selected for further model establishment. After multivariate logistic analysis, the combined model including Mean, age, emphysema, lobulated and size, reached highest AUC of 0.877 (95%CI: 0.830–0.915), accuracy rate of 83.3%, sensitivity of 85.3% and specificity of 80.2% in the training group, followed by radiomics model (AUC: 0.804) and imaging model (AUC: 0.773). A scoring system with a cutoff value greater than 4 points was developed. If the score was larger than 8 points, the possibility of diagnosing malignant IISSPNs could reach at least 92.7%.

Conclusion The combined model demonstrated good diagnostic performance in predicting the malignant potential of IISSPNs. A perfect accuracy rate of 100% can be achieved with a score exceeding 12 points in the user-friendly scoring system.

Keywords Solid pulmonary nodules, Radiomics, Scoring system

*Correspondence:
Xue-Ying Deng
a1145047838@163.com
[1]Department of Radiology, Wenling TCM Hospital Affiliated to Zhejiang Chinese Medical University, Taizhou, Zhejiang 317500, China
[2]Department of Nuclear medicine, Zhejiang Cancer Hospital, Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China
[3]Department of Radiology, Zhejiang Cancer Hospital, Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, Zhejiang 310022, China

Introduction

With the widespread use of high-resolution computed tomography (CT) in screening for pulmonary nodules and increased public health awareness, there has been a significant rise in the identification of incidental solid pulmonary nodules [1]. The current guidelines place important emphasize on the evaluation of the probability of a nodule being malignant, and subsequent treatment depends on the predicted risk of malignancy [2, 3]. The optimal assessment of an individual presenting with a pulmonary nodule would facilitate prompt management of a malignant nodule and reduce unnecessary testing for those with a benign nodule [4]. Approximately 95% of pulmonary nodules detected on CT scans are determined to be benign, most commonly intrapulmonary lymph nodes or granulomas [5]. A biopsy or surgical excision is recommended for high-risk patients, while low-risk patients are monitored through CT followed-up. However, the group with indeterminate-risk poses significant clinical challenges, giving rise to the highest rate of diagnostic error and increasing the risks associated with invasive diagnostic procedures [6]. Highly precise non-invasive risk assessment strategies are imperative in order to decrease mortality rates and overtreatment for those patients.

Risk stratification is a crucial element in the management of indeterminate pulmonary nodules, and various predictive models have been utilized to assess malignant potential [7]. According to current international guidelines, the primary indicators for determining the nature of pulmonary nodules are their size and growth rate [1]. However, assessing and characterizing nodules based solely on their size has certain limitations. In the Brock model, they identified that female sex, older age, family history of lung cancer, emphysema, larger nodule size, upper lobe location, part-solid nodule, lower nodule count, and spiculation were predictions of lung cancer with excellent diagnostic performance, even for nodules smaller than 10 mm in size [8].

Radiomics is a data-driven approach that involves the extraction of numerous quantitative features from medical images using advanced algorithms for image characterization, which can serve as an effective tool in cancer diagnosis, and therapeutic response prediction [9, 10]. Few studies have developed nomograms for predicting the malignant risk of solid pulmonary nodules [11–13], however, their radiomics scores are not easily applicable or reproducible in the clinical practice. A user-friendly scoring system can address this limitation; however, no relevant prior research has been conducted on predicting the malignant risk of pulmonary nodules. Therefore, our study aims to develop a practical scoring system based on risk stratification using measurable features and feasible calculation methods, in conjunction with radiomics and imaging features, for predicting the malignant potential of incidental indeterminate small solid pulmonary nodules (IISSPNs) smaller than 20 mm.

Materials and methods

Patient selection

This retrospective study was approved by the institutional review board, and informed consent was waived. We conducted a retrospective reviewed of the pathological records for all patients who underwent surgical resection due to incidental indeterminate pulmonary nodules between January 2015 and May 2022. The inclusion criteria were as follows: (1) lesions ranged from 5 mm to 20 mm in diameter; (2) no history of malignant tumors; (3) underwent high-resolution chest CT with reconstruction thickness less than 1.5 mm; (4) chest CT scan performed within 1 month before the surgery. The exclusion criteria were as follows: (1) with multiple dominant nodules; (2) poor image quality effected by artifacts; (3) nodules were part-solid or non-solid.

Finally, a total of 360 patients were enrolled in our study (59.9±11.1 years old, ranging from 22 to 87 years old), including 219 (60.8%) males and 141 (39.2%) females. The proportion of malignant IISSPNs was 59.2%, whereas benign nodules accounted for the remaining 40.8%. The most common of malignant IISSPN was lung adenocarcinoma (n=168), followed by squamous cell carcinomas (n=26), neuroendocrine carcinoma (n=7), small cell carcinoma (n=4), undifferentiated carcinoma (n=4), adenosquamous carcinoma (n=3), mucoepidermoid carcinoma (n=1). The most frequent benign IISSPN was inflammatory nodule or proliferative nodule (n=74), followed by granuloma (n=53), hamartoma (n=12), sclerosing alveolar cell tumor (n=4), tuberculoma (n=4).

Chest CT examination

All chest CT scans encompassed the entire thoracic region and were conducted with patients in a supine position. Multiple CT scanners were used in this study as follows: Siemens Somatom Definition Flash 64 (Siemens Medical System, Forchheim, Germany), Optima CT680 Series (GE Medical Healthcare, Milwaukee, Wisconsin). The CT parameters were acquired with a 120 kVp, 60–210 mAs with auto exposure control, using a matrix of 512×512 and pitch values ranging from 0.99 to 1.375. The slice thickness and increment were both set at 5 mm, while the reconstruction layer thickness ranged from 0.75 to 1.25 mm.

Image analysis

The imaging characteristics were acquired by two radiologists who were unaware of the pathology of the nodules. Any disagreements were resolved through consultation with a third radiologist experienced in chest

diseases. Demographic data, including age and gender, were documented. The following imaging information was collected: emphysema, maximum nodule diameter (mm), location (upper vs. middle/lower lobe), subpleural area (whether located within 2 cm of the pleura), shape (round vs. irregular shape), and presence of the lobulated sign, speculated sign, pleural indentation sign.

Radiomics feature extraction

Prior to extraction, pre-processing was necessary to enhance the ability to discriminate between texture features. In the initial stage, data normalization and discretization of grey levels were performed to enhance discrimination between different sets and improve model convergence rate. Subsequently, an eight-level quantization representation was utilized to resample the acquisition area to a specific isotropic resolution (voxel size = $1 \times 1 \times 1$ mm^3) with consistent orientation relative to the plane resolution [14].

The SlicerRadiomics model in the 3D Slicer Radiomics Extension Pack (v.5.0.2 https://www.slicer.org/) was utilized for feature extraction. A semi-automatic 3D segmentation of each pulmonary nodule was performed using the segmentation threshold and seed growing module. The only entity that was meticulously delineated was the tumor, while pleural indentations, spiculated cords around the nodule, and nearby trachea and blood vessels were not included in the delineation process. Radiomics features extracted from 3D nodules were included as follows: the first-order characteristics, grey-level run-length matrix, the grey level co-occurrence matrix, neighborhood grey-tone difference matrix and grey-level size zone matrix [15]. These features could describe the internal or surface texture and morphological features of the lesion. Finally, a total of 1037 features were extracted from lesions. The Mann-Whitney U test or t-test was performed to eliminate features without significant differences, and only those with $p < 0.05$ were included for further analysis. The workflow of our study was listed in Fig. 1.

Feature selection

The entire cohort was randomly divided into training group and validation group at a ratio of 7:3. The training dataset was subjected to dimensional reduction using the least absolute shrinkage and selection operator (LASSO) regression model. The LASSO algorithm debases the dimension of the data by minimizing the residual sum of squares and placing a bound on the sum of the absolute values of the coefficients [16]. The LASSO method was employed with 10-fold cross-validation to perform feature selection, and the minimum λ value was determined to identify the optimal number of selected features.

Development of the scoring system

Three scoring systems were built based on radiomics model, imaging model and their combined model, respectively. To enhance result interpretation and score assignment, the continuous variables were dichotomized for regression analysis, and a receiver operator characteristic (ROC) curve was performed to determine the optimal cutoff values. The selected features after LASSO

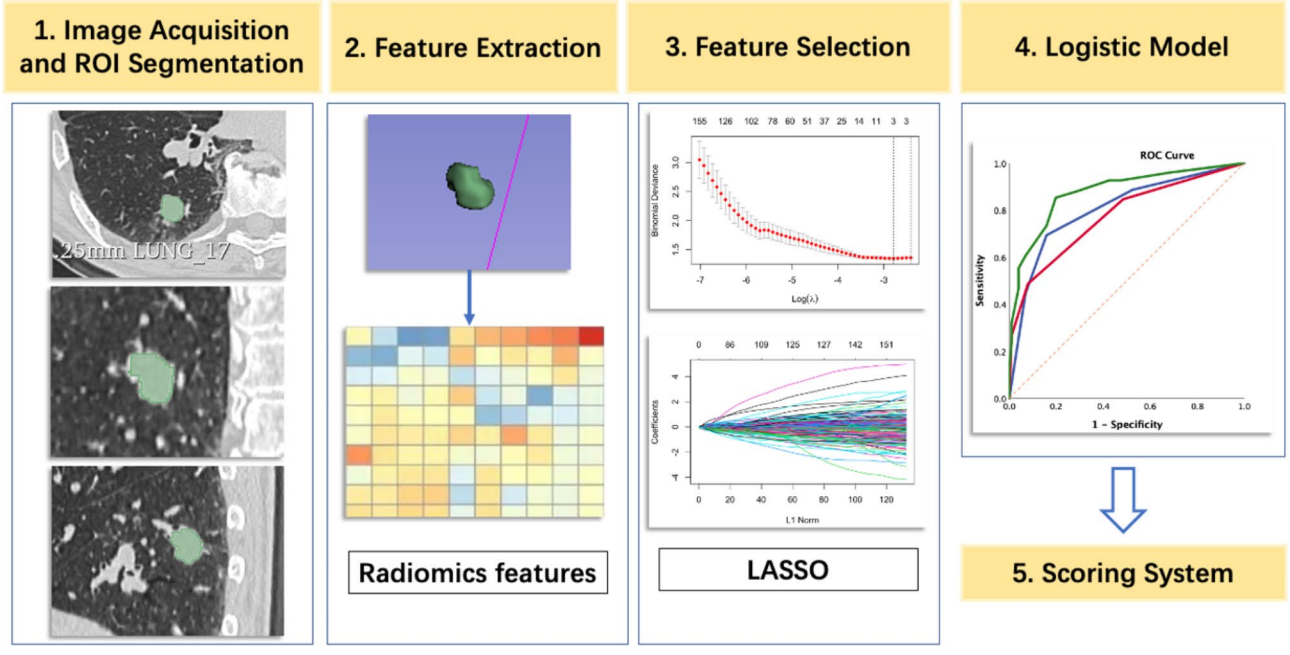

Fig. 1 The framework of our study

algorithm and selected imaging features ($p<0.05$ after univariate analysis) were than included in the multivariable logistic regression analyses to identify the independent differential predictors according to different models. The odds ratio (OR) value and 95% confident interval (CI) for each model were recorded. After conducting binary logistic regression analysis, multiple independent factors were identified and assigned a score of 1, 2, 3, 4, or 5 based on their respective OR values. Subsequently, an ROC curve analysis was conducted to assess the model's diagnostic performance, with threshold, area under the curve (AUC), 95% CI, sensitivity, specificity, and accuracy rate being calculated, according to different models both in training and validation groups. The comparison of ROC curves was conducted using the DeLong test.

Statistical analysis

Quantitative data are commonly presented as either the mean±standard deviation or median (25–75%) values, utilizing the t-test or Mann-Whitney U test based on their distribution. Meanwhile, categorical variables are reported as frequencies (%), using the chi-square test or Fisher exact test. The LASSO algorithm was performed using R software (version 4.1.2; www.R-project.org) using the following R packages: glmnet, foreach, matrix, caret. Univariate and multivariate analysis were analyzed using SPSS 23.0 software (SPSS Inc., Chicago, IL, USA).

Results

Clinical information and imaging characteristics

The demographic information and imaging features are summarized in Table 1. Malignant IISSPNs occurred more frequently in older individuals compared to benign IISSPNs (61.9±10.4 vs. 57.1±11.5 years old, $p<0.001$). They were both more likely to occur in males (64.3% vs. 55.8%, $p=0.103$) and in subpleural area (70.4% vs. 62.5%), with no significant difference found. The maximum diameter of the malignant IISSPNs was generally larger than that of the benign IISSPNs (13.7±3.8 mm vs. 12.1±4.2 mm, $p<0.001$). Emphysema was more often observed in malignant nodules than in benign nodules (31.5% vs. 17.0%), and the difference was significant. Malignant IISSPNs were more likely to exhibit an irregular shape (81.7% vs. 66.0%), a lobulated margin (48.8% vs. 8.2%), and be accompanied by spiculated signs (46.5% vs. 25.9%) and pleural indentations (52.1% vs. 33.3%), compared to benign IISSPNs, and these differences were statistically significant. No significant difference was found in terms of lesion location.

Table 1 General clinical and imaging features between malignant IISSPNs and benign IISSPNs

Variable	Total($n=360$)	Malignant IISSPNs ($n=213$)	Benign IISSPNs ($n=147$)	*P* value
Age(years)	59.9±11.1	61.9±10.4	57.1±11.5	**<0.001**
Gender				0.103
Male	219 (60.8)	137(64.3)	82(55.8)	
Female	141(39.2)	76(35.7)	65(44.2)	
Emphysema	92(25.6)	67(31.5)	25(17.0)	**0.002**
Maximum size(mm)	13.0±4.0	13.7±3.8	12.1±4.2	**<0.001**
Location				0.219
Upper lobe	177(49.2)	99(46.5)	78(53.1)	
Middle/lower lobe	183(50.8)	114(53.5)	69(46.9)	
Irregular shape	271(75.3)	174(81.7)	97(66.0)	**0.001**
Lobulated sign	116(32.2)	104(48.8)	12(8.2)	**<0.001**
Spiculated sign	137(38.1)	99(46.5)	38(25.9)	**<0.001**
Pleural indentation	160(44.4)	111(52.1)	49(33.3)	**<0.001**
Subpleural area	242(67.2)	150(70.4)	92(62.6)	0.119

IISSPNs: incidental indeterminate small solid pulmonary nodules

Feature extraction and selection

In this study, 251 patients were included in the training group, while remaining 109 patients were included in the validation group. No statistically significant differences were observed in either clinical information or imaging features between the training and validation groups (Supplementary Materia Table 1).

In the training group, a total of 1037 features were initially extracted, after eliminating 560 irrelevant statistical radiomics features through univariable analysis, the remaining 477 features were ultimately included in the LASSO algorithm. Finally, only three radiomics features including Minimum (from log-sigma-4-0-mm-3D first order), Mean (from wavelet-HLL first order) and DependenceEntropy (from original gldm) were extracted by the minimum λ value of 0.0613, and minimum standard deviation of -2.415 using the LASSO regression method. (Fig. 2)

Establishment of the logistic model

The outcomes of the multivariable logistic regression analyses are presented in Table 2. After univariable analysis, variables for which $p<0.05$ were included in the multivariable logistic analysis. Prior to further analysis, continuous variables such as DependenceEntropy, Minimum, Mean, age and maximum size were performed to identify the best threshold for dichotomization. The best cutoff values to predict the malignant risk of IISSPNs were as follows: DependenceEntropy>3.87; Minimum ≤-542.32; Mean ≤ -73.911; age>60 years old; maximum size>14.7 mm. Furthermore, subsequent to the completion of the analyses, we ascertained that all of these cutoff values had significant importance ($p<0.05$).

After logistic regression analyses, DependenceEntropy (OR: 2.05, 95%CI: 1.07–3.91) and Mean (OR: 11.60, 95%CI: 6.06–22.22) were identified independent predictors in the radiomics model, while age (OR: 2.66,

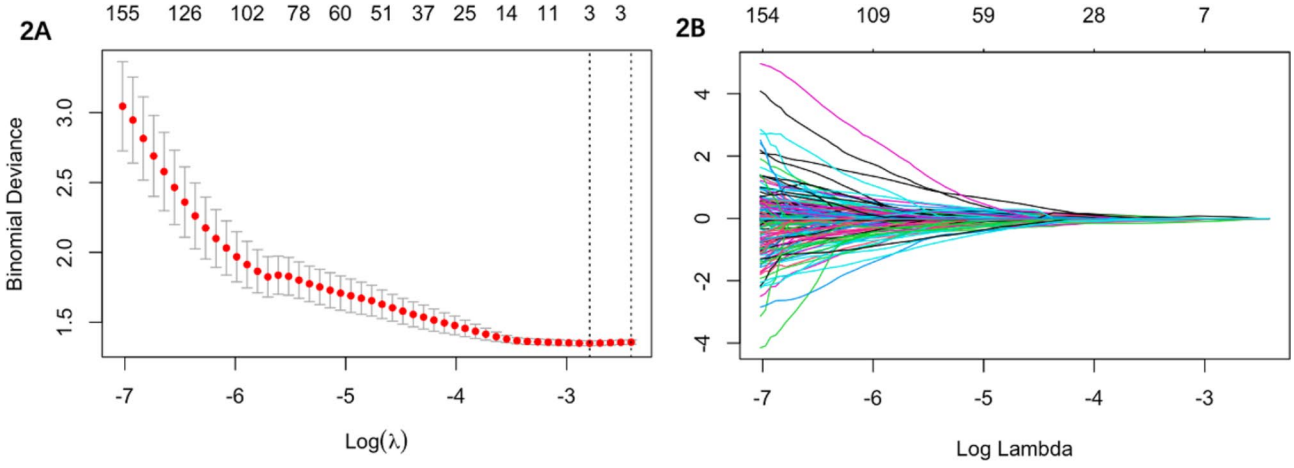

Fig. 2 Radiomics feature selection using the LASSO algorithm. **(A)** The graph depicts the binomial deviance (y-axis) plotted against log(λ). The left dotted line represents the minimum λ value of 0.061, log(λ) = -2.791, while the right dotted line corresponds to the minimum λ standard deviation value of 0.089, log(λ) = -2.415. **(B)** The regularization parameter (λ) was employed for feature reduction. The depicted graph illustrates the variability of coefficients across 477 features as different numbers of features are chosen

Table 2 Multivariable logistic regression analyses in radiomics model, imaging model and their combined model

Variable	Radiomics model		Imaging model		Combined model	
	OR (95%CI)	*p* value	OR (95%CI)	*p* value	OR (95%CI)	*p* value
DependenceEntropy	2.05(1.07–3.91)	**0.030**			0.67(0.28–1.64)	0.388
Minimum	1.33(0.66–2.71)	0.426			0.94(0.41–2.18)	0.888
Mean	11.60(6.06–22.22)	**< 0.001**			12.44(5.85–26.48)	**< 0.001**
Age(years)			2.66(1.42–4.99)	**0.002**	2.45(1.19–5.07)	**0.015**
Emphysema			1.77 (0.85–3.71)	0.130	2.47(1.06–5.75)	**0.036**
Maximum size(mm)			1.71(0.85–3.43)	0.131	2.60(1.04–6.49)	**0.041**
Irregular shape			1.22(0.61–2.40)	0.566	1.13(0.50–2.51)	0.773
Lobulated sign			10.34(4.41–24.28)	**< 0.001**	10.12(3.71–27.58)	**< 0.001**
Spiculated sign			1.66(0.70–3.96)	0.252	2.08(0.75–5.74)	0.158
Pleural indentation			1.05(0.45–2.41)	0.916	0.77(0.29–2.03)	0.595

OR: Odds ratio; 95%CI: 95% confident interval

Table 3 Diagnostic performance of the radiomics model, imaging model and combined model

Models	Data Set	Accuracy (%)	AUC (95%CI)	Sensitivity (%)	Specificity (%)	Cutoff Value
Radiomics model	Training set	75.3	0.804(0.749–0.851)	69.3	84.1	> 1
	Validation set	71.2	0.728(0.635–0.809)	58.7	89.1	> 1
Imaging model	Training set	66.1	0.773(0.717–0.824)	48.7	92.1	> 1
	Validation set	67.0	0.740(0.647–0.819)	49.2	91.3	> 1
Combined model	Training set	83.3	0.877(0.830–0.915)	85.3	80.2	> 4
	Validation set	88.1	0.844(0.762–0.906)	79.4	76.1	> 4

AUC: area under the curve; 95%CI: 95% confident interval

95%CI: 1.42–4.99) and lobulated sign (OR: 10.34, 95%CI: 4.41–24.28) were also found to be independent factors in the imaging model for distinguishing malignant IISSPNs from benign ones. In terms of combined model (radiomics combined with imaging features, we found that Mean (OR: 12.44, 95%CI: 5.85–26.48), lobulated sign (OR: 10.12, 95%CI: 3.71–27.58), maximum size (OR: 2.60, 95%CI: 1.04–6.49), emphysema (OR: 2.47, 95%CI: 1.06–5.75), age (OR: 2.45, 95%CI: 1.19–5.07) were considered as independent predictors for predicting the malignant risk of IISPNs.

Development of the scoring system

The diagnostic performance of the radiomics model, imaging model and combined model is listed in Table 3, and ROC curves are drawn to determine the diagnostic capacity of three models in Fig. 3.

As for radiomics model, according to their OR values, we assigned 1 or 2 points for DependenceEntropy > 3.87,

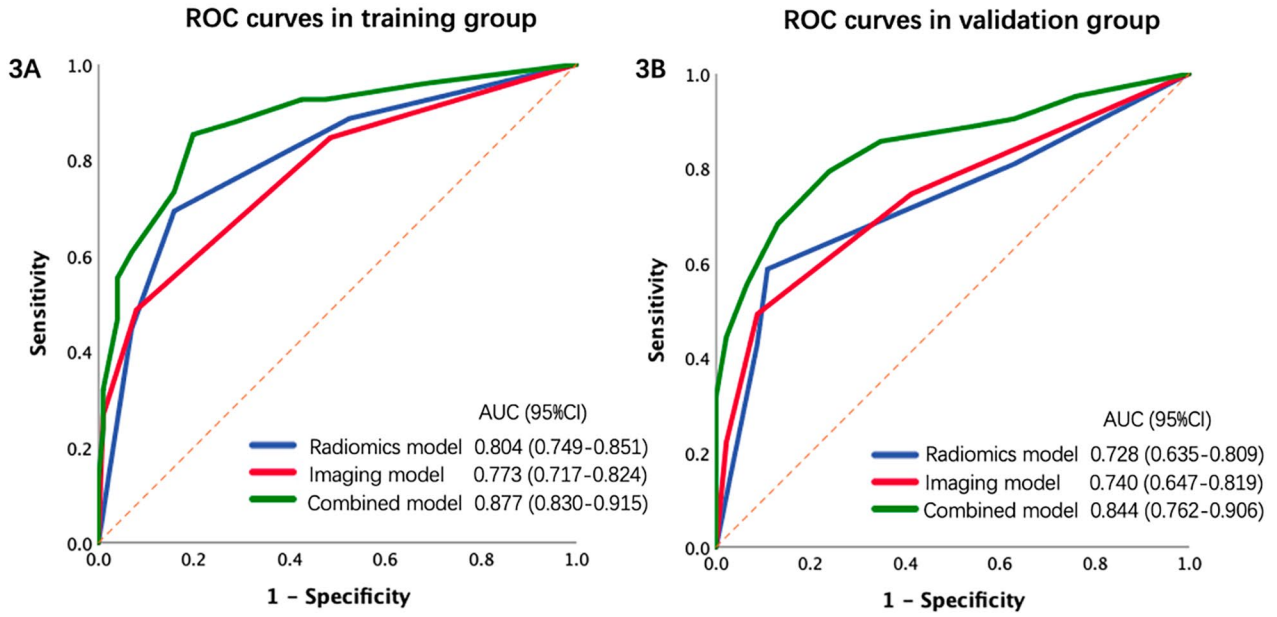

Fig. 3 ROC curves of training and validation groups. **(A)** In the training group, the combined model demonstrated superior diagnostic performance with an AUC of 0.877, followed by radiomics model (AUC: 0.804) and imaging model (AUC: 0.773). **(B)** In the validation group, the combined model achieved the highest performance with an AUC of 0.844, followed by imaging model (AUC: 0.740) and radiomics model (AUC: 0.728)

and Mean≤ -73.911, respectively. ROC curve analysis was performed, and the AUC was 0.804 (95%CI: 0.749–0.851), accuracy rate was 75.3%, sensitivity was 69.3%, specificity was 84.1%, and the cutoff value was larger than 1 point in the training group. The validation group also reached comparable diagnostic performance.

The lobulated sign and age were assigned 2 points and 1 point, respectively, based on their OR values in the imaging model. The AUC was 0.773, 0.740, accuracy rate was 66.1%, 67%, respectively. The cutoff value was both >1 point to reach the best performance.

For the combined model, a score of 5 points was assigned to Mean due to its highest OR value, followed by 4 points for lobulated sign, 3 points for maximum size, 2 points for emphysema, and finally 1 point for age. The combined model achieved the highest AUC of 0.877 (95%CI: 0.830–0.915), with an accuracy of 83.3%, a sensitivity of 85.3%, a specificity of 80.2%, and a cutoff value greater than 4 points in the training group, while achieving an AUC of 0.844 (95%CI: 0.762–0.906) with the highest accuracy of 88.1% in the validation group. Additionally, the possibility of diagnosing malignant IISSPNs could reach 100% when a score was greater than 12 points. Meanwhile, if the score was larger than 4 points but smaller than 9 points, the accuracy rate was 78.4% when distinguishing malignant nodules from benign nodules; if the score was larger than 8 points but smaller than 13 points, the accuracy rate was 92.7%.

The ROC curve of the combined model exhibited a significant and evident correlation with both the imaging model and radiomics model in both training and validation groups ($p<0.05$) as confirmed by Delong test. However, there was no statistically significant difference observed in the ROC curve between the imaging model and radiomics model in both training and validation groups ($p=0.368$, 0.861 respectively).

Discussion

Our research has yielded promising results in the field of predicting the malignant potential of IISPNs pulmonary nodules. It is worth mentioning that we have successfully developed a user-friendly scoring system based on radiomics features and imaging characteristics to differentiate malignant IISPNs from benign IISPNs for the first time. The most compelling result of our study is that the combined model, which includes five predictors, demonstrated superior diagnostic performance, compared to radiomics model or imaging model, based on Delong test. Moreover, an accuracy rate of up to 92.7% can be achieved when the score is greater than 8 points, and a perfect accuracy rate of 100% can be attained with a score exceeding 12 points. This innovative scoring system has allowed us to accurately assess the likelihood of malignancy in IISSPNs nodules. Furthermore, our findings underscore the importance of taking a comprehensive approach to medical diagnosis. Rather than relying on any one type of data or predictor alone, combining multiple sources of information may be key to achieving optimal results.

The incidence of cancer in patients with detected pulmonary nodules ranged from 3.7 to 5.5% [8]. The prevalence of malignancy is a growing concern in the world, with rates varying widely depending on various factors. For example, the size and age were independent predictors in the combined model in our study, which was similar with others [8, 11, 17]. The probability of cancer increases with the size of the tumor, although this relationship is not entirely deterministic. The probability of malignancy in solid nodules measuring 8 mm to 30 mm varies greatly, ranging from very low (<1%) to high (>70%), depending on different risk factors. Even two nodules of equal size may require different management based on their CT appearance. For example, even a 7 mm nodule exhibiting concerning imaging characteristics (such as irregular or spiculated margins, and upper lobe location) raises the likelihood of malignancy to 10%, which may result in more aggressive treatment [4]. In our study, malignant nodules were more likely to present with an irregular shape, a lobulated margin, and be accompanied by spiculated signs and pleural indentations. However, we only found that lobulated sign increased the risk of malignancy after multivariable analysis. Dong et al. [18] found that lobulated shape showed a significant difference in lung adenocarcinoma and tuberculosis. This feature can demonstrate the heterogeneity present with pulmonary nodules, thereby aiding in the differentiation between benign and malignant nodules [19]. The presence of emphysema, was found to be associated with the occurrence of lung cancer, which also been detected in our study. A meta-analysis revealed that emphysema, particularly centrilobular type, was related to a higher odds ratio (OR: 2.3) of developing pulmonary cancer, which also increased with emphysema severity [20].

The management of patients with pulmonary nodules should be considered alongside other individual factors, which is why we developed a user-friendly scoring system. Nomograms have been commonly used in studies on pulmonary nodules to aid in differential diagnosis [12, 13, 17, 21]. These tools allow doctors to make informed decisions about whether further diagnosis or intervention is necessary by quickly calculating risk scores based on multiple variables. However, radiomics typically requires a radiomics score, which is less intuitive as it is derived from imaging data, resulting in limited clinical application. To address this, we proposed a scoring system that assigns different points to various independent predictors based on their respective odds ratios (OR) values. This approach allows for more precise and reliable predictions and is highly convenient for clinical use. For example, Gao et al. [22] built a risk scoring system to assess the prognosis of lung adenocarcinoma patients, while An et al. [23] used a scoring system for predicting gene mutation before treatment. He et al. [24] developed a radiomics-based prognostic scoring system to accurately predict survival outcomes in patients diagnosed with stage IV non-small cell lung cancer undergoing platinum-based chemotherapy. As the first study to predict the malignant potential of isolated incidentally detected IISSPNs, our scoring system demonstrated strong diagnostic performance. In the combined model, a score greater than 12 points can achieve a diagnostic probability of 100% for malignant IISSPNs. Even if the score exceeds 8 points, it still maintains an accuracy rate of at least 92.7%. Our scoring model is relatively more user-friendly and easier to operate, and it may be proficiently applied in clinical practice in the future.

This study has some limitations. Firstly, our study was conducted at a single center and had a retrospective design, resulting in selection bias inevitable. Further study with external validation group is warranted. Secondly, although thin-slice thickness images were included to minimize the influence, the utilization of various CT scans may have an impact on the quality of extracted features. Thirdly, our model only incorporates a subset of variables, with smoking, family history and drinking history being among the omitted factors. In future studies, we aim to incorporate these variables in order to enhance the predictive power of our model.

Conclusions

On the one hand, we established three logistic models, and found that the combined model demonstrated good diagnostic performance in predicting the malignant potential of IISSPNs, superior to the radiomics model or imaging model. On the other hand, a user-friendly scoring system based on radiomics features and imaging characteristics has been developed, which can achieve an accuracy rate of up to 92.7% when the score is greater than 8 points, and a perfect accuracy rate of 100% can be attained with a score exceeding 12 points, when predicting the malignancy of IISSPNs.

Abbreviations

CT	Computed tomography
IISPNs	Incidental indeterminate small solid pulmonary nodules
ROI	Regions of interest
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic curves
OR	Odds ratio
CI	Confidence interval
AUC	Area under the curves

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-024-01413-2.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
Bai-Qiang Qu: Writing original draft, Methodology, Data collection, Yun Wang: Writing editing, Methodology, Supporting, Software; Yue-Peng Pan: Data collection and Feature extraction; Pei-Wei Cao: Data collection and Feature extraction, Software; Xue-Ying Deng: Methodology, Conceptualization, Supervision, Supporting; All authors did literature researches.

Funding
This work was reported by Zhejiang Province Medical, Science and Technology Project, the number is 2021KY091, 2023KY070, 2023KY068, and Zhejiang Provincial Natural Science Foundation, the number is LTGY23H180007.

Data availability
The data was available from the corresponding author with reasonable request.

Declarations

Ethics approval and consent to participate
This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of the Wenling Hospital of Traditional Chinese Medicine. Written informed consent was waived by the institutional review board (Wenling Hospital of Traditional Chinese Medicine).

Consent for publication
NA.

Competing interests
The authors declare no competing interests.

Received: 16 September 2023 / Accepted: 27 August 2024
Published online: 06 September 2024

References

1. Larici AR, Farchione A, Franchi P, Ciliberto M, Cicchetti G, Calandriello L, Del Ciello A, Bonomo L. Lung nodules: size still matters. Eur Respir Rev 2017, 26(146).
2. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, Franks K, Gleeson F, Graham R, Malhotra P, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. Thorax. 2015;70(Suppl 2):ii1–54.
3. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, Mehta AC, Ohno Y, Powell CA, Prokop M, et al. Guidelines for management of Incidental Pulmonary nodules detected on CT images: from the Fleischner Society 2017. Radiology. 2017;284(1):228–43.
4. Mazzone PJ, Lam L. Evaluating the patient with a pulmonary nodule: a review. JAMA. 2022;327(3):264–73.
5. Gould MK, Tang T, Liu IL, Lee J, Zheng C, Danforth KN, Kosco AE, Di Fiore JL, Suh DE. Recent trends in the identification of Incidental Pulmonary nodules. Am J Respir Crit Care Med. 2015;192(10):1208–14.
6. Kammer MN, Lakhani DA, Balar AB, Antic SL, Kussrow AK, Webster RL, Mahapatra S, Barad U, Shah C, Atwater T, et al. Integrated biomarkers for the management of Indeterminate Pulmonary nodules. Am J Respir Crit Care Med. 2021;204(11):1306–16.
7. Kim TJ, Kim CH, Lee HY, Chung MJ, Shin SH, Lee KJ, Lee KS. Management of incidental pulmonary nodules: current strategies and future perspectives. Expert Rev Respir Med. 2020;14(2):173–94.
8. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, Gingras M, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med. 2013;369(10):910–9.
9. Binczyk F, Prazuch W, Bozek P, Polanska J. Radiomics and artificial intelligence in lung cancer screening. Transl Lung Cancer Res. 2021;10(2):1186–99.
10. Gillies RJ, Schabath MB. Radiomics improves Cancer Screening and early detection. Cancer Epidemiol Biomarkers Prev. 2020;29(12):2556–67.
11. Zhang CR, Wang Q, Feng H, Cui YZ, Yu XB, Shi GF. Computed-tomography-based radiomic nomogram for predicting the risk of indeterminate small (5–20 mm) solid pulmonary nodules. Diagn Interv Radiol. 2023;29(2):283–90.
12. Chen C, Geng Q, Song G, Zhang Q, Wang Y, Sun D, Zeng Q, Dai Z, Wang G. A comprehensive nomogram combining CT-based radiomics with clinical features for differentiation of benign and malignant lung subcentimeter solid nodules. Front Oncol. 2023;13:1066360.
13. Huang L, Lin W, Xie D, Yu Y, Cao H, Liao G, Wu S, Yao L, Wang Z, Wang M, et al. Development and validation of a preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule: a machine learning approach, multicenter, diagnostic study. Eur Radiol. 2022;32(3):1983–96.
14. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, et al. 3D slicer as an image computing platform for the quantitative Imaging Network. Magn Reson Imaging. 2012;30(9):1323–41.
15. Sun W, Liu S, Guo J, Liu S, Hao D, Hou F, Wang H, Xu W. A CT-based radiomics nomogram for distinguishing between benign and malignant bone tumours. Cancer Imaging. 2021;21(1):20.
16. Chen HY, Deng XY, Pan Y, Chen JY, Liu YY, Chen WJ, Yang H, Zheng Y, Yang YB, Liu C, et al. Pancreatic serous cystic neoplasms and mucinous cystic neoplasms: Differential diagnosis by combining imaging features and enhanced CT texture analysis. Front Oncol. 2021;11:745001.
17. Liu A, Wang Z, Yang Y, Wang J, Dai X, Wang L, Lu Y, Xue F. Preoperative diagnosis of malignant pulmonary nodules in lung cancer screening with a radiomics nomogram. Cancer Commun (Lond). 2020;40(1):16–24.
18. Dong Q, Wen Q, Li N, Tong J, Li Z, Bao X, Xu J, Li D. Radiomics combined with clinical features in distinguishing non-calcifying tuberculosis granuloma and lung adenocarcinoma in small pulmonary nodules. PeerJ. 2022;10:e14127.
19. Pan X, Wang H, Yu H, Chen Z, Wang Z, Wang L, Chen J. Lung cancer associated with cystic airspaces: CT and pathological features. Transl Cancer Res. 2020;9(6):3960–4.
20. Yang X, Wisselink HJ, Vliegenthart R, Heuvelmans MA, Groen HJM, Vonder M, Dorrius MD, de Bock GH. Association between chest CT-defined Emphysema and Lung Cancer: a systematic review and Meta-analysis. Radiology. 2022;304(2):322–30.
21. Feng B, Chen X, Chen Y, Liu K, Li K, Liu X, Yao N, Li Z, Li R, Zhang C, et al. Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. Eur J Radiol. 2020;128:109022.
22. Gao C, Zhuang J, Li H, Liu C, Zhou C, Liu L, Feng F, Sun C. Gene signatures of 6-methyladenine regulators in women with lung adenocarcinoma and development of a risk scoring system: a retrospective study using the cancer genome atlas database. Aging. 2021;13(3):3957–68.
23. An W, Fan W, Zhong F, Wang B, Wang S, Gan T, Tian S, Liao M. Development and validation of a Concise Prediction Scoring System for Asian Lung Cancer patients with EGFR Mutation before Treatment. Technol Cancer Res Treat. 2022;21:15330338221078732.
24. He L, Li Z, Chen X, Huang Y, Yan L, Liang C, Liu Z. A radiomics prognostic scoring system for predicting progression-free survival in patients with stage IV non-small cell lung cancer treated with platinum-based chemotherapy. Chin J Cancer Res. 2021;33(5):592–605.

Publisher's note