Prediction of malignancy for solitary pulmonary nodules based on imaging, clinical characteristics and tumor marker levels

Hongjun Hou, Shui Yu, Zushan Xu, Hongsheng Zhang, Jie Liu and Wenjun Zhang

Objective To establish a prediction model of malignancy for solitary pulmonary nodules (SPNs) on the basis of imaging, clinical characteristics and tumor marker levels.

Methods Totally, 341 cases of SPNs were enrolled in this retrospective study, in which 70% were selected as the training group (n=238) and the rest 30% as the verification group (n=103). The imaging, clinical characteristics and tumor marker levels of patients with benign and malignant SPNs were compared. Influencing factors were identified using multivariate logistic regression analysis. The model was assessed by the area under the curve (AUC) of the receiver operating characteristic curve.

Results Differences were evident between patients with benign and malignant SPNs in age, gender, smoking history, carcinoembryonic antigen (CEA), neuron-specific enolase, nodule location, edge smoothing, spiculation, lobulation, vascular convergence sign, air bronchogram, ground-glass opacity, vacuole sign and calcification (all \( P<0.05 \)). Influencing factors for malignancy included age, gender, nodule location, spiculation, vacuole sign and CEA (all \( P<0.05 \)). The established model was as follows: \( Y=−5.368+0.055\times \text{age}+1.012\times \text{gender}+2.164\times \text{vacuole sign (yes=1, no=0)}−0.054\times \text{CEA}. \)

Conclusions The prediction model established in our study exhibits better accuracy and internal stability in predicting the probability of malignancy for SPNs.

Keywords: computerized tomography, prediction model, solitary pulmonary nodule, tumor markers

Introduction A solitary pulmonary nodule (SPN) refers to a single, round radiologic opacity that is not more than 3 cm in its maximum diameter and at least moderately well marginated, without distal atelectasis, pleural effusion, local lymph node enlargement and peripheral satellite lesions (Patel et al., 2013; Truong et al., 2014). The SPN frequently encountered on chest imaging may be solid or subsolid in attenuation. Subsolid nodules containing a component with ground-glass attenuation are highly likely to develop into premalignant or malignant lesions after the presence of 3–6 months (Naidich et al., 2013; Borghesi et al., 2020). Eight large trials of lung cancer screening suggested that the prevalence of at least one SPN ranged from 8 to 51%, and that of malignancy in patients with an SPN varied from 1.1 to 12.0% (Wahidi et al., 2007). With the development of imaging technology and growing interest in lung cancer screening, the detection of SPNs has increased markedly. How to characterize and treat SPNs becomes a major concern for clinicians.

Multidetector computerized tomography (MDCT) plays an important role in assessing the morphological characteristics and nodal growth on serial images. It improves the specificity and sensitivity in the detection of pulmonary nodules by reducing misregistration artifacts and increasing spatial and contrast resolution, which provides more accurate characterization of nodules (Truong et al., 2014; Snoeckx et al., 2018). As reported in the various guidelines for pulmonary nodules, the nodule size and morphological characteristics are closely associated with the probability of malignancy (MacMahon et al., 2005; Callister et al., 2015; MacMahon et al., 2017). Additionally, the clinical context should not be neglected when the probability of malignancy is assessed (Larici et al., 2017). There is evidence suggesting that the probability of malignancy for SPNs may be increased in patients with advanced
age, smoking and history of extrathoracic malignant neoplasm (Kikano et al., 2015). Although multiple prediction models of malignancy for SPNs have been established currently, including Veterans Affairs model, Mayo Clinic model and Brock model, serum tumor markers are almost missing in these models (Swensen et al., 1997; Gould et al., 2007; Chung et al., 2018).

It is well-known that the detection of serum tumor markers contributes to the screening and early diagnosis of lung cancer. By measuring the tumor marker levels in patients with pulmonary nodules, Li et al., found that the levels of serum cytokeratin fragment antigen 21-1 (CYFRA21-1) and carcinoembryonic antigen (CEA) were increased significantly in patients with malignant nodules when compared with those with benign nodules, suggesting a potential effect of serum tumor markers in determining benign and malignant pulmonary nodules (Li et al., 2017). In this study, a prediction model of malignancy for SPNs was developed to assist the clinicians to effectively diagnose and treat SPNs based on the imaging features obtained from MDCT, clinical characteristics and serum tumor markers.

**Methods**

**Study population**

Between 1 July 2017 and 31 December 2019, totally 341 cases of SPNs who underwent surgical excision or puncture biopsy at Weihai Central Hospital were enrolled in this retrospective study. This study was performed on the basis of the principle of Helsinki Declaration and was approved by the Institutional Review Board of Weihai Central Hospital (approval no.: WHSZXYKYLL-2020-03).

The included patients must meet all the following requirements, including: (1) the nodule diameter ≤3.0 cm, and no pulmonary atelectasis, pneumonia, satellite lesions and local lymph node enlargement; (2) able to cooperate better and almost keep the same breath-holding depth at each scanning; (3) complete pathological results and (4) the nature of nodule confirmed by pathological findings and clinical characteristics. Exclusion criteria were as follows: (1) patients with lung cancer-related symptoms, such as irritable cough, bloody sputum and chest pain, or those with local lymph node enlargement; (2) patients with obvious damage or severe insufficiency of important organs like the heart, liver and kidney; (3) pregnant women; (4) patients with a metal stent or internal fixator in the chest-back area, or those with common contraindications to iodinated contrast agents; (5) patients with SPNs extremely close to the heart margin or diaphragm and (6) patients taking other trial drugs 1 month before enrollment or participating in other trials.

**Image acquisition**

The thin-section computed tomography (CT) scanning was performed on the patient’s chest using the GE BrightSpeed 16-layer MDCT scanner. The lung window (window level: -400HU; window width: 500HU) and mediastinal window (window level: 50HU; window width: 500HU) were set, respectively, and the section thickness was 1.25 mm. Additionally, 50 mL of iodinated contrast agent (300 mg I/mL) was injected in superficial veins of the forearm using a high-pressure injector, with a flow rate of 4 mL/s. The lesions were dynamically scanned at 30 s and 50 s after the injection of iodinated contrast agent, and the prolonged time of scanning was 5.6 s.

For each patient, two imaging specialists were independently responsible for reading the film and carefully recording the CT features of pulmonary nodules, including: (1) nodule location; (2) nodule size, namely the maximum diameter of nodule measured in the lung window; (3) edge features, such as smoothing, spiculation, lobulation, pleural indentation, vascular convergence sign, etc. and (4) internal characteristics like cavity and calcification. If there existed conflicts about the description of nodules, the third imaging specialist would involve in determination.

**Measurement of serum tumor markers**

Before surgery and drug use, 5 mL of peripheral venous blood from fasting patients in the morning was drawn and centrifuged to collect serum samples, which were stored at −20°C. The levels of serum tumor markers CEA, neuron-specific enolase (NSE), CYFRA21-1, carbohydrate antigen 125 (CA125) and CA199 were detected using electrochemiluminescence assay, and the kits were provided by Roche Diagnostics Ltd., Shanghai. All the operations were implemented based on the kit instructions.

**Statistical analysis**

SAS software (version 9.4, SAS Institute Inc., North Carolina, USA.) was used for statistical analysis. The data with normal and abnormal distributions were described as the mean ± standard deviations (X ± s) using Student’s t-test and the median and quartile [M (Q25, Q75)] by Mann–Whitney U test, respectively. Categorical data were expressed as the number of cases and its proportion [n (%)] using the chi-square test or Fisher’s exact test. For investigating the influencing factors for benign and malignant SPNs, the variables with significant difference in univariate analysis were enrolled into a multivariate logistic regression model through stepwise regression. A prediction model of malignancy for SPNs was established, which was assessed by the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. A significant difference was shown at \( P < 0.05 \).

**Results**

**Baseline information of study population**

Totally, 341 patients with SPNs were enrolled in this study between 1 July 2017 and 31 December 2019, with the mean age of (59.53 ± 8.97) years. There were 172 males aged (60.02 ± 8.38) years and 169 females aged (59.04 ± 9.53) years. Among them, 162 cases suffered from benign nodules, whereas 179 from malignant nodules.
Of 341 patients, 70% were randomly selected as the training group \((n = 238)\) and the rest 30% were as the verification group \((n = 103)\). The baseline information of training and verification groups is compared in Table 1. It was found that the differences were not pronounced between two groups in age, gender, smoking history, history of malignant neoplasm and complication of other diseases \((P > 0.05)\), which suggested that the data of the verification group could be used to internally verify the prediction model established by the data of training group.

**Clinical characteristics and tumor marker levels of patients with benign and malignant solitary pulmonary nodule**

In total, 238 cases in the training group were at the age of \((59.47 \pm 9.00)\) years, including 113 males and 125 females. Among them, there were 113 cases of benign SPNs and 125 cases of malignant SPNs.

The clinical characteristics and serum tumor marker levels of patients with benign and malignant SPNs are listed in Table 2. The results showed that the age of patients with malignant SPNs was older than those with benign SPNs \((P = 0.001)\). Compared with those with benign SPNs, the patients with malignant SPNs had a higher proportion of females \((P = 0.001)\) and a lower proportion of smoking history \((P = 0.002)\). In terms of serum tumor markers, the patients with malignant SPNs had a lower CEA level \((P = 0.003)\) and a higher NSE level compared with those with benign SPNs \((P = 0.033)\) and a higher NSE level compared with those with benign SPNs \((P = 0.001)\) and a higher proportion of smoking history \((P = 0.002)\).

**Imaging features of patients with benign and malignant solitary pulmonary nodule**

The imaging features of patients with benign and malignant SPNs are compared in Table 3. It was shown that the differences were apparent between the patients with benign and malignant SPNs in the nodule location, presence or absence of edge smoothing, spiculation, lobulation, vascular convergence sign, air bronchogram, ground-glass opacity, vacuole sign and calcification (all \(P < 0.05\)).

**Multivariate logistic regression analysis of benign and malignant solitary pulmonary nodule**

As shown in Table 4, multivariate Logistic regression analysis exhibited that the age [odds ratio (OR), 1.056; 95% confidence interval (CI), 1.016–1.097; \(P = 0.005\)], gender (OR, 2.750; 95% CI, 1.451–5.212; \(P = 0.002\)), nodule location (OR, 3.677; 95% CI, 1.558–8.675; \(P = 0.003\)), spiculation (OR, 3.347; 95% CI, 1.628–6.880; \(P = 0.001\)) and vacuole sign (OR, 8.706; 95% CI, 3.765–20.134; \(P < 0.001\)) were the independent risk factors for malignant SPNs, whereas serum CEA level (OR, 0.948; 95% CI, 0.905–0.993; \(P = 0.024\)) was a protective factor. Based on this, a prediction model for the malignancy of SPNs was established, namely \(Y = -5.568 + 0.055 \times \text{age} + 1.012 \times \text{gender (female = 1, male = 0)} + 1.302 \times \text{nodule location} \left(\text{right upper lobe} = 1, \text{others} = 0\right) + 1.208 \times \text{spiculation (yes = 1, no = 0)} + 2.164 \times \text{vacuole sign (yes = 1, no = 0)} - 0.054 \times \text{CEA}; \)

\[
P = \dfrac{e^Y}{1 + e^Y}.
\]

The larger the \(P\) value, the greater the risk of malignancy.

**Diagnostic value and validation of the model**

The diagnostic values of the age, CEA level, and models with CEA and without CEA are compared in Table 5. It can be observed that the model with CEA had a sensitivity of 64.80% and specificity of 84.96%, with the cut-off value of 0.61, and its AUC was 0.818 (95% CI, 0.736–0.865), significantly larger than 0.793 of the model without CEA (95% CI, 0.736–0.842; \(P = 0.029\)), 0.621 of the age alone (95% CI, 0.556–0.683; \(P < 0.001\)), and 0.580 of the CEA level alone (95% CI, 0.515–0.643; \(P < 0.001\)). Figure 1 represents the ROC curves of each model. The internal verification of the model was performed using the data of the verification group. No significant difference was presented between the predictive number and actual number of the model with CEA in predicting malignant SNPs \((P = 0.384, \text{Table } 6)\), suggesting that this prediction model had better internal stability.

### Table 1  Baseline information of training and verification groups, \(n(\%)\)

| Variables                        | Number of cases | Training group \((n = 238)\) | Verification group \((n = 103)\) | \(Z/\chi^2\) | \(P\) value |
|----------------------------------|----------------|-------------------------------|---------------------------------|--------------|------------|
| Age, years, \(M (Q_{25}, Q_{75})\) | 60 (55, 65)    | 59 (53, 66)                   |                                 | −0.474       | 0.635      |
| Gender                           |                |                               |                                 | 2.763        | 0.096      |
| Male                             | 172            | 113 (47.48)                   | 59 (57.28)                      |              |            |
| Female                           | 169            | 125 (52.52)                   | 44 (42.72)                      |              |            |
| Smoking history                  |                |                               |                                 | 3.537        | 0.060      |
| No                               | 230            | 168 (70.59)                   | 62 (60.19)                      |              |            |
| Yes                              | 111            | 70 (29.41)                    | 41 (39.81)                      |              |            |
| History of malignant neoplasm    |                |                               |                                 |              |            |
| No                               | 337            | 236 (99.16)                   | 101 (98.06)                     |              |            |
| Yes                              | 4              | 2 (0.84)                      | 2 (1.94)                        |              |            |
| Comorbidity                      |                |                               |                                 | 0.524        | 0.469      |
| No                               | 239            | 164 (68.91)                   | 75 (72.82)                      |              |            |
| Yes                              | 102            | 74 (31.09)                    | 28 (27.18)                      |              |            |

*\(\chi^2\) represented the data were analyzed using Fisher’s exact test.
A risk prediction model of malignancy for SPNs Hou et al.

Assessment of the model calibration

The model calibration was assessed based on the training and verification groups. Hosmer–Lemeshow goodness of fit test showed a better calibration in the training group ($\chi^2 = 3.600; P = 0.892$) and the verification group ($\chi^2 = 4.375; P = 0.822$). In addition, the calibration curves also indicated that the model had a better calibration (Fig. 2).

Discussion

CT has long been recognized as the most sensitive and widespread imaging method in the detection and diagnosis of SPNs, especially the MDCT. For depicting pulmonary nodules, MDCT is more sensitive than single-detector CT due to its higher contrast and spatial resolution (Truong et al., 2014). Although $^{18}$F-fluorodeoxyglucose PET imaging of focal lesions plus CT scan is thought to

Table 2  Clinical characteristics and tumor marker levels of patients with benign and malignant solitary pulmonary nodules, n (%)

| Variables                        | Number of cases | Benign SPNs (n=113) | Malignant SPNs (n=125) | Z/χ²  | P     |
|----------------------------------|-----------------|----------------------|------------------------|-------|-------|
| Age, years, M (Q₂₅, Q₇₅)        | 113             | 58.00 (54.00, 63.00) | 62.00 (56.00, 67.00)   | -3.215| 0.001 |
| Gender                           |                 |                      |                        | 10.304| 0.001 |
| Male                             | 66 (58.41)      | 47 (37.60)           |                        |       |       |
| Female                           | 47 (41.59)      | 78 (62.40)           |                        |       |       |
| Smoking history                   | 168             | 69 (61.06)           | 99 (79.20)             | 9.405 | 0.002 |
| Yes                              | 70 (48.94)      | 26 (20.80)           |                        |       |       |
| History of malignant neoplasm    |                 |                      |                        | 4.999 |       |
| No                               | 236             | 113 (100.00)         | 123 (98.40)            |       |       |
| Yes                              | 2               | 0 (0)                | 2 (1.60)               |       |       |
| Comorbidity                       | 164             | 83 (73.45)           | 81 (64.80)             | 2.073 | 0.150 |
| Yes                              | 74              | 30 (26.55)           | 44 (35.20)             |       |       |
| CEA, μg/L, M (Q₂₅, Q₇₅)         |                 | 2.50 (1.40, 3.19)    | 1.91 (1.36, 2.88)      | 2.130 | 0.033 |
| NSE, μg/L, M (Q₂₅, Q₇₅)         |                 | 11.20 (10.10, 12.60) | 12.07 (10.40, 14.10)   | -2.195| 0.028 |
| CYFRA21-1, ng/mL, M (Q₂₅, Q₇₅)   |                 | 10.33 (6.89, 12.70)  | 9.52 (8.51, 11.50)     | 1.346 | 0.178 |
| CA199, U/mL, M (Q₂₅, Q₇₅)       |                 | 8.80 (5.20, 10.10)   | 6.90 (4.90, 9.10)      | 1.238 | 0.216 |

'–' represented the data were analyzed using Fisher’s exact test.

CA, carbohydrate antigen; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin fragment antigen 21-1; NSE, neuron-specific enolase; SPNs, solitary pulmonary nodules.

Table 3  Imaging features of patients with benign and malignant solitary pulmonary nodules, n (%)

| Variables                        | Number of cases | Benign SPNs (n=113) | Malignant SPNs (n=125) | Z/χ²  | P     |
|----------------------------------|-----------------|----------------------|------------------------|-------|-------|
| Nodule size, cm, M (Q₂₅, Q₇₅)    |                 | 1.90 (1.10, 2.80)    | 1.60 (1.10, 2.30)      | 1.627 | 0.104 |
| Nodule location                   |                 |                      |                        | 10.482| 0.033 |
| Left upper lobe                   | 69              | 36 (31.86)           | 33 (26.40)             |       |       |
| Left lower lobe                   | 41              | 21 (18.58)           | 20 (16.00)             |       |       |
| Right upper lobe                  | 66              | 23 (20.35)           | 43 (34.40)             |       |       |
| Right middle lobe                 | 19              | 14 (12.39)           | 5 (4.00)               |       |       |
| Right lower lobe                  | 43              | 19 (16.81)           | 24 (19.20)             |       |       |
| Smoothing                         |                 |                      |                        | 34.655| <0.001|
| No                               | 201             | 70 (69.91)           | 122 (97.60)            |       |       |
| Yes                              | 37              | 34 (30.09)           | 3 (2.40)               |       |       |
| Spiculation                       |                 |                      |                        | 11.438| <0.001|
| No                               | 66              | 43 (38.05)           | 23 (18.40)             |       |       |
| Yes                              | 172             | 70 (61.95)           | 102 (81.60)            |       |       |
| Lobulation                        |                 |                      |                        | 23.655| <0.001|
| No                               | 195             | 107 (94.69)          | 88 (70.40)             |       |       |
| Yes                              | 43              | 6 (5.31)             | 37 (29.60)             |       |       |
| Vascular convergence sign         |                 |                      |                        | 65.500| <0.001|
| No                               | 81              | 68 (60.18)           | 13 (10.40)             |       |       |
| Yes                              | 157             | 45 (39.82)           | 112 (89.60)            |       |       |
| Pleural indentation               |                 |                      |                        | 1.399 | 0.237 |
| No                               | 166             | 83 (73.45)           | 83 (65.40)             |       |       |
| Yes                              | 72              | 30 (26.55)           | 42 (33.60)             |       |       |
| Air bronchogram                   |                 |                      |                        | 11.306| <0.001|
| No                               | 207             | 107 (94.69)          | 100 (80.00)            |       |       |
| Yes                              | 31              | 6 (5.31)             | 25 (20.00)             |       |       |
| Ground-glass opacity              |                 |                      |                        | 74.997| <0.001|
| No                               | 154             | 105 (62.92)          | 49 (39.20)             |       |       |
| Yes                              | 84              | 8 (70.8)             | 76 (60.80)             |       |       |
| Vacuole sign                      |                 |                      |                        | 28.974| <0.001|
| No                               | 174             | 101 (89.38)          | 73 (58.40)             |       |       |
| Yes                              | 64              | 12 (10.62)           | 52 (41.60)             |       |       |
| Calcification                     |                 |                      |                        | 8.724 | 0.003 |
| No                               | 227             | 103 (91.15)          | 124 (99.20)            |       |       |
| Yes                              | 11              | 10 (8.85)            | 1 (0.80)               |       |       |

SPNs, solitary pulmonary nodules.
be a well-established technique for assessing ambiguous lesions more than 1 cm, it has the major limitations of unreliable imaging for nodules less than 1 cm, low sensitivity to ground-class nodules and a high cost (Herder et al., 2004; Nomori et al., 2004; Pašnik et al., 2017). In the present study, MDCT was used to describe the imaging features of 341 patients with SPNs, and a risk prediction model for SPNs was developed based on the patients’ imaging, clinical characteristics and serum tumor marker levels. Multivariate Logistic regression model showed that age, gender, nodule location, spiculation and vacuole sign were the independent risk factors for malignant SPNs, whereas serum CEA level was a protective factor. Based on this, a risk prediction model for SPNs was established: \[ Y = -5.368 + 0.055 \times \text{age} + 1.012 \times \text{gender} (\text{female} = 1, \text{male} = 0) + 1.302 \times \text{nodule location} (\text{right upper lobe} = 1, \text{others} = 0) + 1.208 \times \text{spiculation} (\text{yes} = 1, \text{no} = 0) + 2.164 \times \text{vacuole sign} (\text{yes} = 1, \text{no} = 0) - 0.054 \times \text{CEA}; \]

\[ P = \frac{e^Y}{1 + e^Y}, \]

which was confirmed to have better accuracy and internal stability in predicting the probability of malignancy.

Generally, the older the patients, the higher the risk of developing tumors. The time of tumorigenic factors acting on the body is prolonged with age, leading to an increased probability of developing tumors, and the lung cancer is no exception (Toomes et al., 1983; Erasmus et al., 2000). Yankelevitz et al. (1999) found that the patients with SPNs aged over 40 years were related to a higher risk of lung cancer. Our results demonstrated that age was independent risk factors for malignant SPNs, and was involved in the establishment of the model, which was supported by several mathematical models including the Veterans Affairs model and Mayo Clinic model (Swensen et al., 1997; Gould et al., 2007). However, gender was not included in the above models. According to the presence or absence of spiculation, two prediction models for malignant SPNs developed in a prospective study suggested that the female was a risk factor for malignancy, with ORs of 1.76 and 1.82, respectively (McWilliams et al., 2013). In the present study, gender was found to be a significant risk factor for malignant SPNs, and the probability of malignancy in females was 2.750 folds higher than that in males, slightly higher than the OR of 2.103 in another study (Wang et al., 2018a). Previous evidence showed that the common pathological pattern of non-small cell lung cancer in most females, especially young females, was lung adenocarcinoma, and the proportion of lung adenocarcinoma could come up to 80% in Asian females with lung cancer (Subramanian et al., 2010; Hu and Li, 2012).

In the present study, the nodule location (right upper lobe), presence of spiculation and vacuole sign as the significant risk factors for malignant SPNs were all enrolled in the prediction model. The evidence has affirmed that numerous primary malignant pulmonary nodules are placed in the upper lobes, especially in the right, although 2/3 of metastatic nodules have an impact on the lower lobes (Khan et al., 2011). The risk of malignancy in the right upper lobe was the highest, approximately accounting for 45% of all malignant nodules (Horeweg et al., 2013). It may be explained by increased airflow into the right upper lobe during the initial inspiration and thereby increased exposure to inhaled carcinogens (Horeweg et al., 2013; Cruickshank et al., 2019). In addition, malignant nodules usually present with irregular, spiculated or lobulated margins because of malignant cells spreading into the pulmonary interstitium, whereas benign nodules

---

**Table 4 Multivariate logistic regression analysis of benign and malignant solitary pulmonary nodules**

| Variables                               | \( \beta \) | S.E.  | Wald   | \( P \) value | OR     | 95% CI  |
|-----------------------------------------|------------|------|--------|--------------|--------|--------|
| Constant                                | −5.368     | 1.279| 17.607 | <0.001       | 1      | 1.016  |
| Age, years                              | 0.055      | 0.020| 7.833  | 0.005        | 1.056  | 1.016  |
| Gender (female vs. male)                | 1.012      | 0.326| 9.622  | 0.002        | 2.750  | 1.451  |

**Table 5 Diagnostic values of each model**

| Model                        | AUC     | 95% CI        | Sensitivity | Specificity | Youden index | Cut-off | Z     | \( P \) value |
|------------------------------|---------|---------------|-------------|-------------|--------------|---------|-------|---------------|
| CEA, \( \mu g/L \)           | 0.580   | 0.515–0.643   | 72.00       | 47.79       | 0.20         | 2.65    | 5.565 | <0.001        |
| Age                          | 0.821   | 0.566–0.883   | 36.80       | 84.96       | 0.22         | 64      | 5.148 | <0.001        |
| The model without CEA        | 0.793   | 0.738–0.842   | 68.00       | 77.88       | 0.46         | 0.56   | 2.189 | 0.029         |
| The model with CEA           | 0.818   | 0.763–0.869   | 64.80       | 84.96       | 0.50         | 0.01   |       |               |

CEA, carcino-embryonic antigen; CI, confidence interval; OR, odds ratio; SE, standard error.
tend to have smooth, rounded edges (Choromańska and Macura, 2012). Spiculation is pathologically associated with desmoplastic reactions and may also be caused by the infiltration of interstitial planes and lymphatics by tumor (Khan et al., 2011). Enormous evidence have demonstrated a strong association between spiculated margins and malignant pulmonary nodules (Ost and Gould 2012; Hu et al., 2016; Ferreira et al., 2018). Vacuole sign, unlike air bronchogram, is the area of low attenuation owing to small patent air containing bronchi in nodules. The presence of vacuole sign is also reported to be correlated with malignant pulmonary nodules (Snoeckx et al., 2018; Yu et al., 2020). In most studies, nodule size is found to be important in the prediction of benign and malignant SPNs (Yu et al., 2020; Wang et al., 2018a,b). However, in the present study, it was not identified as a parameter of the model, which might be associated with study populations and sample size.

A prediction model for malignant SPNs that had been affirmed to have better diagnostic accuracy, internal stability and calibration was developed in the present study, with several potential implications. First, unlike most prediction models for malignant pulmonary nodules, the serum tumor marker was included in this model to further improve the diagnostic accuracy. This model was demonstrated to have the AUC of 0.818 and a better internal stability using the data of verification group. Second, the data from this model were easy to be obtained in clinic, which may help the clinicians to rapidly evaluate the probability of malignancy. Third, it was of great significance to enhance the recognition of patients at high risk of malignancy through this prediction model. Although the high proportion of nonsmokers (67.4%) in our study was similar as several Chinese studies, it was not representative for other countries (Gould et al., 2013; Truong et al., 2014; Yu et al., 2020). Additionally, the establishment of the model was based on a retrospective study with relatively small sample sizes; the model was not simply enough for clinician use and lack of external validation. In the future, external validation would be required to establish the robustness and reliability of the model.
validation will be further implemented in large-scale, prospective studies.

**Conclusion**

The influencing factors for malignant SPNs include age, gender, nodule location, spiculation, vacuole sign and CEA level, based on which a prediction model is established. It is demonstrated to have better accuracy and internal stability in predicting the probability of malignancy for SPNs.

**Acknowledgements**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This study was performed based on the principle of Helsinki Declaration and was approved by the Institutional Review Board of Weihai Central Hospital (approval No.: WHSZYYYKL-2020-03).

**Conflicts of interest**

There are no conflicts of interest.

**References**

Borgiess A, Michelin S, Goleni S, Scrimenti A, Maroldi R (2020). What’s new on quantitative CT analysis as a tool to predict growth in persistent pulmonary subsolid nodules? A literature review. *Diagnoses (Basel)* 10:E35.

Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Drafjan J, et al.; British Thoracic Society Pulmonary Nodule Guideline Development Group; British Thoracic Society Standards of Care Committee (2015). British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 70 (Suppl 2):i1–i54.

Choromatsa A, Macura KJ (2012). Evaluation of solitary pulmonary nodule detected during computed tomography examination. *Pol J Radiol* 77:22–34.

Chung K, Mets OM, Gerke PK, Canep B, Drafjan J, et al.; British Thoracic Society Pulmonary Nodule Guideline Development Group; British Thoracic Society Standards of Care Committee (2015). British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 70 (Suppl 2):i1–i54.

Ferreira JR Jr, Oliveira MC, de Azevedo-Marques PM (2018). Characterization of pulmonary nodules based on features of margin sharpness and texture. *J Digit Imaging* 31:451–463.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 1: radiologic characteristics and imaging of subsolid nodules. *Radiology* 212:59–66.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 2: differential diagnosis and clinical management. *Radiology* 212:67–73.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 3: follow-up and surgical management. *Radiology* 212:75–79.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 4: differential diagnosis and clinical management. *Radiology* 212:81–85.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 5: follow-up and surgical management. *Radiology* 212:87–90.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 6: differential diagnosis and clinical management. *Radiology* 212:91–95.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 7: follow-up and surgical management. *Radiology* 212:97–100.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 8: differential diagnosis and clinical management. *Radiology* 212:101–104.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 9: follow-up and surgical management. *Radiology* 212:105–108.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 10: differential diagnosis and clinical management. *Radiology* 212:109–112.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 11: follow-up and surgical management. *Radiology* 212:113–116.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 12: differential diagnosis and clinical management. *Radiology* 212:117–120.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 13: follow-up and surgical management. *Radiology* 212:121–124.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 14: differential diagnosis and clinical management. *Radiology* 212:125–127.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 15: follow-up and surgical management. *Radiology* 212:128–131.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 16: differential diagnosis and clinical management. *Radiology* 212:132–135.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 17: follow-up and surgical management. *Radiology* 212:136–139.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 18: differential diagnosis and clinical management. *Radiology* 212:140–143.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 19: follow-up and surgical management. *Radiology* 212:144–147.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 20: differential diagnosis and clinical management. *Radiology* 212:148–151.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 21: follow-up and surgical management. *Radiology* 212:152–155.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 22: differential diagnosis and clinical management. *Radiology* 212:156–159.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 23: follow-up and surgical management. *Radiology* 212:160–163.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 24: differential diagnosis and clinical management. *Radiology* 212:164–167.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 25: follow-up and surgical management. *Radiology* 212:168–171.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 26: differential diagnosis and clinical management. *Radiology* 212:172–175.

Gould MK, McAdams HP, Connolly JE (2000). Solitary pulmonary nodules: part 27: follow-up and surgical management. *Radiology* 212:176–179.