



A random forest algorithm predicting model combining intraoperative frozen section analysis and clinical features guides surgical strategy for peripheral solitary pulmonary nodules

Liqiang Qian^{1#}, Yinjie Zhou^{2#}, Wanqin Zeng³, Xiaoke Chen¹, Zhengping Ding¹, Yujia Shen³, Yifeng Qian⁴, Davide Tosi⁵, Mario Silva⁶, Yuchen Han⁷, Xiaolong Fu³

¹Shanghai Lung Cancer Center, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China; ²Department of Thoracic Surgery, Hwa Mei Hospital, University of Chinese Academy of Science (Ningbo No. 2 Hospital), Ningbo, China; ³Department of Radiation Oncology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China; ⁴National Clinical Research Center for Oral Disease, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University, Shanghai, China; ⁵Thoracic Surgery and Lung Transplant Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; ⁶Scienze Radiologiche, Department of Medicine and Surgery (DiMeC), University of Parma, Parma, Italy; ⁷Department of Pathology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

Contributions: (I) Conception and design: X Fu, L Qian; (II) Administrative support: X Fu; (III) Provision of study materials or patients: Y Han, Y Zhou, Z Ding; (IV) Collection and assembly of data: W Zeng, X Chen; (V) Data analysis and interpretation: Y Shen, Y Qian; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally and should be considered as co-first authors.

Correspondence to: Xiaolong Fu, MD. Department of Radiation Oncology, Shanghai Chest Hospital, Shanghai Jiao Tong University, 241 West Huaihai Road, Shanghai 200030, China. Email: xlfu1964@hotmail.com; Yuchen Han, MD. Department of Pathology, Shanghai Chest Hospital, Shanghai Jiao Tong University, 241 West Huaihai Road, Shanghai 200030, China. Email: ychan@cmu.edu.cn.

Background: Intraoperative frozen section (FS) analysis has been used to guide the extent of resection in patients with solitary pulmonary nodules (SPNs), but its accuracy varies greatly among different hospitals. Artificial intelligence (AI) and multidimensional data technology are developing rapidly these years, meanwhile, surgeons need better methods to guide the surgical strategy of SPNs. We established predicting models combining FS results with multidimensional perioperative clinical features using logistic regression analysis and the random forest (RF) algorithm to get more accurate extent of SPN resection.

Methods: Patients with peripheral SPNs who underwent FS-guided surgical resection at the Shanghai Chest Hospital (January 2017–December 2018) were retrospectively examined (N=3,089). The accuracy of intraoperative FS-guided resection extent was analyzed and used as Model 1. The clinical features (sex, age, CT features, tumor markers, smoking history, lesion size and nodule location) of patients were collected, and Models 2 and 3 were established using logistic regression and RF algorithms to combine the FS with clinical features. We confirmed the performance of these models in an external validation cohort of 117 patients from Hwa Mei Hospital, University of Chinese Academy of Science (Ningbo No. 2 Hospital). We compared the effectiveness in classifying low/high-risk groups of SPN among them.

Results: The accuracy of FS analysis was 61.3%. Model 3 exhibited the best diagnostic accuracy and had an area under the curve of 0.903 in the internal validation cohort and 0.919 in the external validation cohort. The calibration plots and net reclassification index (NRI) of Model 3 also exhibited significantly better performance than the other models. Improved diagnostic accuracy was observed in both internal and external validation cohort.

Conclusions: Using an RF algorithm, clinical characteristics can be combined with intraoperative FS analysis to significantly improve intraoperative judgment accuracy for low- and high-risk tumors, and may serve as a reliable complementary method when FS evaluation is equivocal, improving the accuracy of the extent of surgical resection.

Keywords: Solitary pulmonary nodule (SPN); frozen section (FS); surgical resection; diagnostic accuracy; random forest (RF)

Submitted Mar 28, 2022. Accepted for publication Jun 16, 2022.

doi: 10.21037/tlcr-22-395

View this article at: <https://dx.doi.org/10.21037/tlcr-22-395>

Introduction

Most solitary pulmonary nodules (SPNs) are identified in the early stage through pathological diagnosis and are potentially curable. However, the accurate diagnosis of SPNs is clinically challenging because lesions may represent inflammation, infection, benign lung tumors, or other non-malignant issues (1). Many SPNs with ground glass opacity (GGO) components are diagnosed as lung adenocarcinomas or precancerous lesions, such as adenomatous atypical hyperplasia (AAH), adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), or invasive adenocarcinoma (IAC) (2).

The extent of surgical resection for SPNs varies according to the diagnosis. For malignant lung tumors classified as high risk, the standardized surgical method involves lobectomy and systemic node dissection because of the probability of postoperative recurrence and metastasis (3). However, following the publication of the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification in 2011 (4), several studies, such as Zhang *et al.* (5), have reported early-stage lung adenocarcinomas (e.g., AAH, AIS, and MIA) are associated with good prognosis, and sublobar resection without lymphadenectomy is currently considered a more appropriate surgical procedure. The same applies to benign tumors and some low-grade malignant tumors (such as carcinoid tumors) (6), classified as low risk. Therefore, the classification of SPNs determines the extent of surgical resection, which is crucial for optimized planning of tailored surgical approach aiming to minimal invasiveness while maintaining radical intent.

Nonetheless, it is difficult to diagnose SPNs preoperatively because of the significant uncertainties with the application of computed tomography (CT), bronchoscopy, and needle biopsy (7,8). Frozen sections (FS) of specimens resected during surgery have become the primary diagnostic modality for SPNs. FS are used to determine both the benign or malignant nature of SPNs and extent of tumor infiltration for low-risk or high-risk

malignant tumors. As they are used to guide surgeons in determining the extent of surgical resection, there is a critical need for achieving high diagnostic accuracy when using FS.

Whether FS accurately determine the properties and infiltration degree of SPNs remains controversial. Liu *et al.* suggested intraoperative FS accurately determine the degree of tumor infiltration and guide the resection strategy in patients with lung adenocarcinoma (9). However, other studies have found a certain error rate in determining the tumor infiltration degree of lung adenocarcinoma solely based on FS compared with the final pathology (FP) (10,11). Better predictions of the final pathological outcome of lung adenocarcinoma have been achieved by combining FS results with tumor diameter (12). Furthermore, SPNs may represent other pathological diagnoses other than lung adenocarcinoma, rendering FS-guided diagnosis more challenging. Currently, no large-scale studies investigating the accuracy of intraoperative FS in determining SPN properties and guiding surgical resection exist.

In recent years, artificial intelligence and machine learning (ML) have been widely used in various fields, including medicine (13). Artificial intelligence and ML algorithms analyze large volumes of data by learning a decisional process, which can be continuously refined for improved performance (14). The random forest (RF) algorithm is an important ML algorithm. It's essentially an ensemble learning algorithm based on bagging. Its basic principle is to combine multiple weak classifiers, and the final results are voted or averaged, so that the results of the overall model have high accuracy and better generalization. The clinical features, such as sex, age, CT features, tumor markers, smoking history, lesion size and nodule location are very suitable to be defined as "weak classifiers" in the RF algorithm.

This real-world study aimed to evaluate the accuracy of the extent of SPN resection under intraoperative FS guidance using logistic regression analysis and the RF algorithm to establish a model combining FS results with multidimensional perioperative clinical information. We

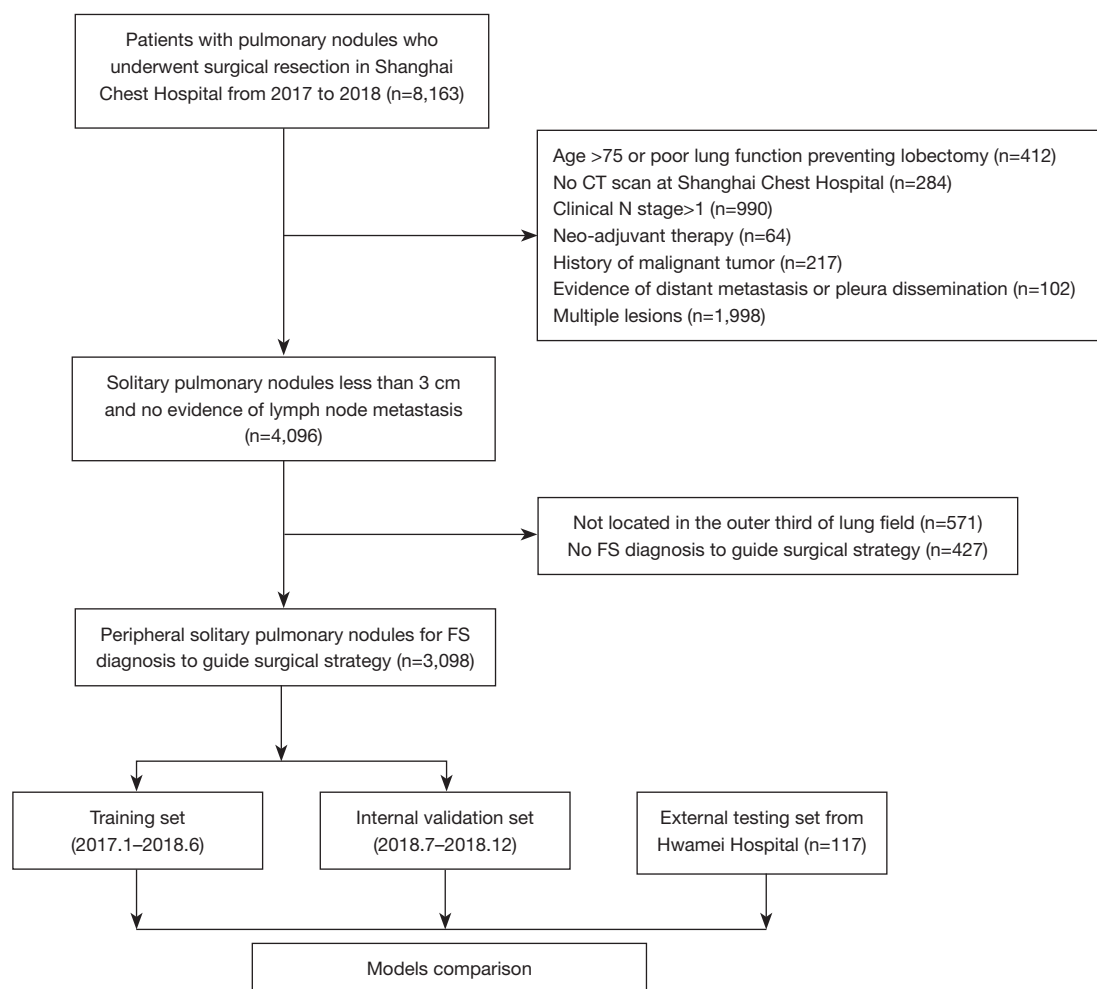


Figure 1 Flowchart of patient inclusion. CT, computed tomography; FS, frozen section.

verified whether this model could improve the accuracy of intraoperative SPN classification. We present the following article in accordance with the STARD reporting checklist (available at <https://tldr.amegroups.com/article/view/10.21037/tlcr-22-395/rc>).

Methods

Study cohort and data collection

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and approved by the Committee of Medical Ethics of Shanghai Chest Hospital (approval number KS21002, 2021-2) and Hwa Mei Hospital, University of Chinese Academy of Science (Ningbo No. 2 Hospital, approval number YJ-NBEY-KY-2021-140-01, 2021-9). Informed consent was waived

because of the retrospective nature of the study. We retrospectively analyzed all peripheral SPN (located at the outer 1/3 of lung field) resections performed under the guidance of intraoperative FS at the Shanghai Chest Hospital between January 2017 and December 2018. An external cohort from Hwa Mei Hospital, University of Chinese Academy of Science (Ningbo No. 2 Hospital) between January 2017 and December 2018 were also collected and used for external validation. The research design of this study and exclusion criteria are shown in *Figure 1*. Preoperative tests (contrast-enhanced chest CT, abdominal CT or ultrasonography, brain magnetic resonance imaging or CT, and radionuclide bone scan for most patients and positron emission tomography or CT for the rest) were performed to assess the clinical stage of the lesion. Clinicopathologic data, such as sex, age at surgery,

CT features (GGO component and pleura indentation), presence of tumor markers, smoking history, lesion size measured in fresh specimens, nodule location, resection type, FS diagnosis, and FP, were collected.

CT scans and tumor markers

CT imaging and tumor marker assessments were performed approximately 1 week (6.8 ± 3.2 days) before surgery. Most chest CT scans were contrast-enhanced (some GGOs were not). CT scans in Shanghai Chest Hospital were obtained using Brilliance iCT and Brilliance 64 CT scanners (Philips Healthcare, Eindhoven, Netherlands). Each nodule was reviewed twice by two radiologists (YLM and TGY) with 15 and 10 years of experience, respectively, and CT features were distinguished based on the presence of GGO (nodule with/without GGO component) and pleural indentation. The standard values of tumor markers were as follows: carcinoembryonic antigen, 0–5 ng/mL; carbohydrate antigen 19-9, 0–5 ng/mL; cytokeratin 19 fragment, 0–1.5 ng/mL; neuron-specific enolase, 0–25 ng/mL; and cancer antigen 125, 0–35 U/mL.

Evaluation of FS and final pathology findings

After the tumors were removed via sublobar resection, pathologists immediately performed FS diagnosis of the specimens, and if the lesion was diagnosed as adenocarcinoma, the presence of AAH, AIS, MIA, and IAC were determined. After FS diagnosis, the specimens were immersed in 10% neutral buffered formalin and embedded in paraffin. All FS diagnoses were compared with the final pathologic diagnoses of the corresponding permanent paraffin sections. The pathological diagnoses were made according to the 2015 World Health Organization classification for lung tumors and 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society Classification.

The pathologies of pulmonary nodules were divided into the following seven categories: AAH, AIS, MIA, IAC, other types of malignant tumors (squamous cell carcinoma, large-cell carcinoma, and lymphoepithelioid-like carcinoma), low-grade malignant tumors (carcinoid), and benign findings (pulmonary hamartoma, adenoma, granuloma, aspergilloma, tuberculosis, and inflammation). They were then divided into two groups: high-risk (IAC and other types of malignant tumors) and low-risk (AAH, AIS, MIA,

low-grade malignant tumors, and benign tumors) groups.

Compared with FP, the concordance of the FS results was defined as follows: “correct” (consistent with FP), “underestimated”, “overestimated”, “error” (misjudged between benign and malignant), and “equivocal” (or deferred).

Surgical procedures

During surgery, sublobar resection (including wedge resection and segmentectomy) was first performed, followed by FS pathological examination. If the FS pathological result indicated a high-risk classification, subsequent lobectomy and lymph node dissection were performed. However, if the FS pathological result was equivocal or deferred, the surgical team determined the extent of resection based on experience.

Model establishment and statistical analysis

The dataset of Shanghai Chest Hospital was divided into two cohorts by the date of surgery as follows: (I) training cohort including patients who underwent surgery from January 2017 to June 2018; and (II) internal validation cohort including patients who underwent surgery from July 2018 to December 2018. An independent dataset from Hwa Mei Hospital, University of Chinese Academy of Science (Ningbo No. 2 Hospital) as an external validation cohort (15). We use cross validation for calibration.

First, we established a univariate logistic regression model based on FS alone for the diagnosis of high- or low-risk groups (Model 1), which was then combined with patient characteristics, including age, sex, smoking history, maximum SPN diameter in CT, with/without GGO component, pleural indentation, and tumor marker results, to derive a multivariable logistic regression model (Model 2). Finally, the RF binary classification models were trained using the same features in Model 2 (Model 3).

The classification performance of the abovementioned models was evaluated using confusion matrix analysis, which included accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). Youden’s index was applied to calculate the sensitivity and specificity. These models were also evaluated using a receive operating characteristic curve (ROC) to calculate the area under the curve (AUC) and 95% confidence interval (CI). The threshold of AUC is set to 0.5, which means when the output probability of a case is larger than 0.5, then the case is considered as high-risk group. Delong Test was

applied for comparing AUC values of the three models. We also generated calibration plots and determined the net reclassification index (NRI). Results were compared in both the internal and external validation cohorts.

All statistical analyses, model building, and model evaluation were performed in R using the caret package (version 3.5.2; R Foundation for Statistical Computing, Vienna, Austria; <http://www.r-project.org>). Statistical significance was defined as a two-sided P value <0.05.

Results

Patient data

A total of 8,163 patients with pulmonary nodules underwent surgical resection at Shanghai Chest Hospital, from January 2017 to December 2018. Data were analyzed for the 3,098 patients with peripheral SPNs ≤ 3 cm who met the inclusion criteria, and their clinicopathological characteristics are summarized in *Table 1*.

Lymph node metastasis was observed in IAC 1 cm in CT screening (both N1 and N2) in diameter and squamous carcinoma 1.5 cm (N1) in diameter, whereas it was not observed in patients with AAH/AIS/MIA or other malignant tumors <1 cm who underwent systemic lymphadenectomy or lymph node sampling. Results for tumor markers (including carcinoembryonic antigen, carbohydrate antigen 19-9, cytokeratin 19 fragment, neuron-specific enolase, and cancer antigen 125) were considered in both the training cohort (n=2,059) and internal validation cohort (n=963). The clinicopathological characteristics of the external cohort are summarized in *Table S1*.

FS and surgical procedure accuracy

The comparison of FS and FP results is shown in *Table 2*. The FS concordance compared with FP was: AAH, 81.3%; AIS, 34.3%; MIA, 8.8%; IAC, 77.2%; other types of malignancy, 88.7%; low-grade malignancy, 40%; and benign, 98.4%. FS results compared with FP results were as follows (stratified by pathological type): correct, 1,898 (61.3%); underestimated, 54 (1.7%); overestimated, 100 (3.2%); error, 12 (0.4%); and equivocal, 1,034 (33.4%). FS results compared with FP results were classified as follows (stratified by high or low-risk): correct, 2,022 (65.3%); underestimated, 19 (0.6%); overestimated, 23 (0.7%); and equivocal, 1,034 (33.4%).

Tumor size, MIA pathology, and GGO components were

identified as risk factors for incorrect FS determination in the univariate and multivariate regression analyses (*Table 3*). The accuracy of the extent of surgical resection was as follows: correct surgical extent, 81.2% (n=2,516), and incorrect surgical extent, 18.8% (n=582). Of the 1,034 patients with equivocal FS results, the extent of resection was correct in 494 (47.8%) and incorrect proportion is 52.2% (540 patients, 277 too large and 263 too small).

Models

As shown in *Figure 2A*, the AUC for Model 1 was 0.633 in the internal validation cohort (95% CI: 0.603–0.662), whereas those for Models 2 and 3 were 0.889 (0.869–0.909) and 0.903 (0.884–0.922), respectively. Comparison among the three models revealed the AUC of Model 3 was significantly larger than that of Models 1 and 2 (P<0.001 and P=0.012, respectively). The classification performance of the models was also evaluated using the NRI and calibration plots (the number of replications is 1,000). Comparison of the NRI between Models 2 and 3 is presented in *Figure 2B* and calibration plots in the internal validation cohort are presented in *Figure 2C–2E*. The NRI between these two models was 0.06 (0.03–0.10), indicating that Model 3 exhibited significantly better reclassification performance. The AUC of the three models in the external validation cohort is shown in *Figure 3A*, the AUC for Model 1 was 0.639 (95% CI: 0.553–0.726), whereas those for Models 2 and 3 were 0.889 (0.830–0.948) and 0.919 (0.871–0.967), respectively. Comparison among the three models revealed the AUC of Model 3 was significantly larger than that of Model 1 (P<0.001), with no significant difference between Models 2 and 3 (P=0.196). Comparison of the NRI between Models 2 and 3 is presented in *Figure 3B* and the calibration plots for the external validation cohort are presented in *Figure 3C–3E*. The NRI comparison between Models 2 and 3 was 0.10 (0.01–0.27), also indicating Model 3 exhibited significantly better reclassification performance.

We conducted confusion matrix analysis and assessed the accuracy, sensitivity, specificity, NPV, and PPV of the three models, and their comparison is shown in *Table 4*. Model 3 exhibited the best diagnostic accuracy, with >80% accuracy, sensitivity, specificity, PPV, and NPV in both the internal and external validation cohorts.

Discussion

In this research, we aimed to establish an ML model to

Table 1 Clinicopathologic characteristics of patients included in the study (N=3,098)

Characteristics	Total (N=3,098)	AAH (n=16)	AIS (n=432)	MIA (n=634)	IAC (n=1,385)	Other infiltrative malignancies (n=62)	Low-grade malignancies (n=5)	Benign (n=564)	P
Age (mean ± SD) (years)	54.57±10.96	55.3±9.84	49.3±11.4	50.7±11.7	58.1±9.14	63.4±7.0	63.2±4.9	53.2±10.7	<0.001
Sex, n (%)									<0.001
Male	1,237 (39.9)	3 (18.8)	138 (31.9)	176 (27.8)	576 (41.6)	53 (85.5)	2 (40.0)	289 (51.2)	
Female	1,861 (60.1)	13 (81.2)	294 (68.1)	458 (72.2)	809 (58.4)	9 (14.5)	3 (60.0)	275 (48.8)	
Surgical methods, n (%)									<0.001
Wedge resection	1,104 (35.6)	11 (68.8)	267 (61.8)	284 (44.8)	62 (4.5)	1 (1.6)	2 (40.0)	477 (84.6)	
Segmentectomy	536 (17.3)	3 (18.8)	133 (30.8)	215 (33.9)	95 (6.9)	2 (3.2)	2 (40.0)	86 (15.2)	
Lobectomy	1,458 (47.1)	2 (12.6)	32 (7.4)	135 (21.3)	1,228 (88.6)	59 (95.2)	1 (20.0)	1 (0.2)	
Location of tumor, n (%)									0.002
RUL	1,106 (35.7)	9 (56.3)	174 (40.3)	236 (37.2)	513 (37.0)	18 (29.0)	1 (20.0)	155 (27.5)	
RML	84 (2.7)	1 (6.3)	9 (2.1)	21 (3.3)	18 (1.3)	0	0	35 (6.2)	
RLL	541 (17.5)	0	50 (11.6)	106 (16.7)	238 (17.2)	7 (11.3)	2 (40.0)	138 (24.5)	
LUL	860 (25.8)	5 (31.2)	141 (32.6)	178 (28.1)	391 (28.3)	23 (37.1)	0	122 (21.6)	
LLL	507 (16.3)	1 (6.2)	58 (13.4)	93 (14.7)	225 (16.2)	14 (22.6)	2 (40.0)	114 (20.2)	
Maximum diameter of tumor, n (%)									<0.001
≤1 cm	1,341 (43.3)	13 (81.3)	407 (94.2)	478 (75.4)	151 (10.9)	3 (4.9)	3 (60.0)	286 (50.7)	
1 < d ≤2 cm	1,226 (39.6)	2 (12.5)	23 (5.3)	151 (23.8)	801 (57.8)	26 (41.9)	1 (20.0)	222 (39.4)	
2 < d ≤3 cm	531 (17.1)	1 (6.2)	2 (0.5)	5 (0.8)	433 (31.3)	33 (53.2)	1 (20.0)	56 (9.9)	
Lymph node situation, n (%)									0.957
N0	3,007 (97.1)	16 (100.0)	432 (100.0)	634 (100.0)	1,301 (93.9)	55 (88.7)	5 (100.0)	564 (100.0)	
N1	31 (1.0)	0	0	0	29 (2.1)	2 (3.2)	0	0	
N2	60 (1.9)	0	0	0	55 (4.0)	5 (8.1)	0	0	
CT imaging, n (%)									<0.001
GGO component									<0.001
With	2,378 (76.8)	16 (100.0)	432 (100.0)	634 (100.0)	1,017 (73.4)	1 (1.6)	1 (20.0)	277 (49.1)	
Without	720 (23.2)	0	0	0	368 (26.6)	61 (98.4)	4 (80.0)	287 (50.9)	
Pleural indentation									<0.001
Yes	1,135 (36.6)	0	88 (20.4)	182 (28.7)	826 (59.6)	38 (61.3)	0	1 (0.2)	
No	1,963 (63.4)	16 (100.0)	344 (79.6)	452 (71.3)	559 (40.4)	24 (38.7)	5 (100.0)	563 (99.8)	
Smoking history, n (%)									0.001
Yes	239 (7.7)	2 (12.5)	12 (2.8)	26 (4.1)	119 (8.6)	44 (71.0)	0	36 (6.4)	
No	2,859 (92.3)	14 (87.5)	420 (97.2)	608 (95.9)	1,266 (91.4)	18 (29.0)	5 (100.0)	528 (93.6)	

Table 1 (continued)

Table 1 (continued)

Characteristics	Total (N=3,098)	AAH (n=16)	AIS (n=432)	MIA (n=634)	IAC (n=1,385)	Other infiltrative malignancies (n=62)	Low-grade malignancies (n=5)	Benign (n=564)	P
Tumor biomarkers (mean ± SD)									
CEA	2.76±7.9	2±1.18	1.77±1.51	1.88±1.26	3.66±11.5	4.37±6.12	2.29±1.27	2.66±1.74	0.438
CA19-9	2.48±1.17	2.24±1.24	2.28±0.98	2.38±1.09	2.54±1.11	3.11±1.27	2.83±0.87	2.57±1.44	<0.001
CYFRA21-1	0.85±0.85	0.98±0.80	0.812±0.50	0.847±0.82	0.861±0.98	1.06±0.71	0.68±0.11	0.843±0.79	0.711
NSE	18.14±6.68	16.5±6.27	17.9±6.62	17.7±6.44	18.5±7.05	17.00±4.61	17±3.62	18±6.28	0.685
CA125	12.06±15.4	11±5.53	13.6±35.2	12.1±10.3	11.5±7.83	11±4.54	10.5±3.64	12.3±9.05	0.433

Minimum 1 cm adenocarcinoma with lymph node metastasis (N1, N2), 1.5 cm squamous carcinoma with lymph node metastasis (N1). AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; SD, standard deviation; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; CT, computed tomography; GGO, ground glass opacity; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; CYFRA21-1, cytokeratin 19 fragment antigen21-1; NSE, neuron-specific enolase; CA125, cancer antigen 125.

Table 2 Comparison of frozen section and final pathology results

Frozen section results	Final pathology results, n (%)						
	AAH (n=16)	AIS (n=432)	MIA (n=634)	IAC (n=1,385)	Other types of malignancy (n=62)	Low-grade malignancy (n=5)	Benign (n=564)
AAH	13 (81.3*)	13 (3.0)	3 (0.5)	4 (0.3)	0	0	0
AIS	0	148 (34.3*)	23 (3.6)	2 (0.1)	0	0	0
MIA	0	77 (17.8)	56 (8.8*)	9 (0.6)	0	0	0
IAC	1 (6.2)	12 (2.8)	10 (1.6)	1,068 (77.2*)	1 (1.6)	0	0
Other types of malignancy	0	0	0	0	55 (88.7*)	1 (20.0)	0
Low-grade malignancy	0	0	0	0	0	2 (40.0*)	1 (0.2)
Benign	0	4 (0.9)	1 (0.2)	2 (0.1)	2 (3.2)	1 (20.0)	555 (98.4*)
Equivocal	2 (12.5)	178 (41.2)	541 (85.3)	300 (21.7)	4 (6.5)	1 (20.0)	8 (1.4)

*, indicates the frozen section accuracy for each type of solitary pulmonary nodule. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma.

determine the invasion status of SPNs to aid surgeons in decision-making regarding the extent of surgical resection and lymphadenectomy. CT screening assists with identifying some early-stage lung cancers, particularly those associated with favorable histology (16), and increasing interest in sublobar resection has been shown to preserve lung function, to reduce perioperative morbidity, and to provide a chance of resection for a subsequent primary lung cancer (17,18). To date, the optimal extent of surgical resection and lymphadenectomy remains controversial. Sublobar resection without lymph node dissection may be the preferred surgical procedure for some low-grade

malignancies and early-stage lung adenocarcinomas (5,6). However, “spread through air spaces (19)” and lymph node metastases may still be present for some lung malignancies with smaller diameters (≤ 1 cm) (20) (similar to this study) and require lobectomy and lymphadenectomy. It was recently revealed that it is inappropriate to decide on surgical strategies solely based on imaging performance, because many GGO-predominant nodules may also be IACs, and the extent of infiltration cannot be determined based on the amount of GGO component (7).

Although FS may represent a better choice for guiding the surgical strategy, its small specimen volume makes FS-

Table 3 Univariable and multivariable analyses of factors contributing to incorrect frozen section diagnosis

Variable	Univariate analysis		Multivariate analysis	
	OR (95% CI)	P	OR (95% CI)	P
Age	0.97 (0.97, 0.98)	<0.001	1.01 (1.00, 1.02)	0.246
Maximum diameter	0.29 (0.25, 0.33)	<0.001	0.38 (0.30, 0.47)	<0.001
Sex				
Male	Reference		Reference	
Female	1.80 (1.54, 2.11)	<0.001	1.13 (0.91, 1.40)	0.272
Location				
LUL	Reference		Reference	
LLL	0.97 (0.77, 1.23)	0.819	1.30 (0.95, 1.76)	0.096
RUL	1.15 (0.95, 1.38)	0.147	1.20 (0.94, 1.52)	0.138
RML	0.95 (0.58, 1.51)	0.825	1.14 (0.55, 2.35)	0.725
RLL	0.81 (0.64, 1.02)	0.075	0.95 (0.70, 1.30)	0.769
Pathology				
AAH	Reference		Reference	
AIS	3.40 (1.08, 14.99)	0.059	3.99 (1.24, 17.80)	0.035
MIA	28.77 (9.05, 127.41)	<0.001	43.50 (13.36, 195.34)	<0.001
IAC	1.29 (0.41, 5.64)	0.696	5.15 (1.58, 23.14)	0.013
Other types of cancer	0.46 (0.11, 2.42)	0.320	8.30 (1.72, 47.26)	0.010
Low-grade malignancy	1.08 (0.05, 11.69)	0.950	3.55 (0.13, 49.48)	0.367
Benign	0.06 (0.02, 0.31)	<0.001	0.16 (0.04, 0.79)	0.013
Pleural indentation				
No	Reference		Reference	
Yes	0.77 (0.66, 0.90)	0.001	0.86 (0.69, 1.06)	0.154
GGO component				
No	Reference		Reference	
Yes	16.26 (11.55, 23.72)	<0.001	4.27 (2.85, 6.62)	<0.001
History of smoking				
No	Reference		Reference	
Yes	0.51 (0.37, 0.70)	<0.001	0.95 (0.61, 1.46)	0.807

OR, odds ratio; CI, confidence interval; LUL, left upper lobe; LLL, left lower lobe; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; GGO, ground glass opacity.

guided determination more challenging. It is also difficult to interpret lung tissue FS because of severely distorted architecture, ice crystal formation, and the complete collapse of the alveolar spaces during cryosection. This

issue is of particular concern for the determination of MIA when stromal invasion is ≤ 5 mm. MIA leads to a diagnosis of IAC, and neglecting the invasive component leads to a diagnosis of AIS. In this study, it was also found that lesions

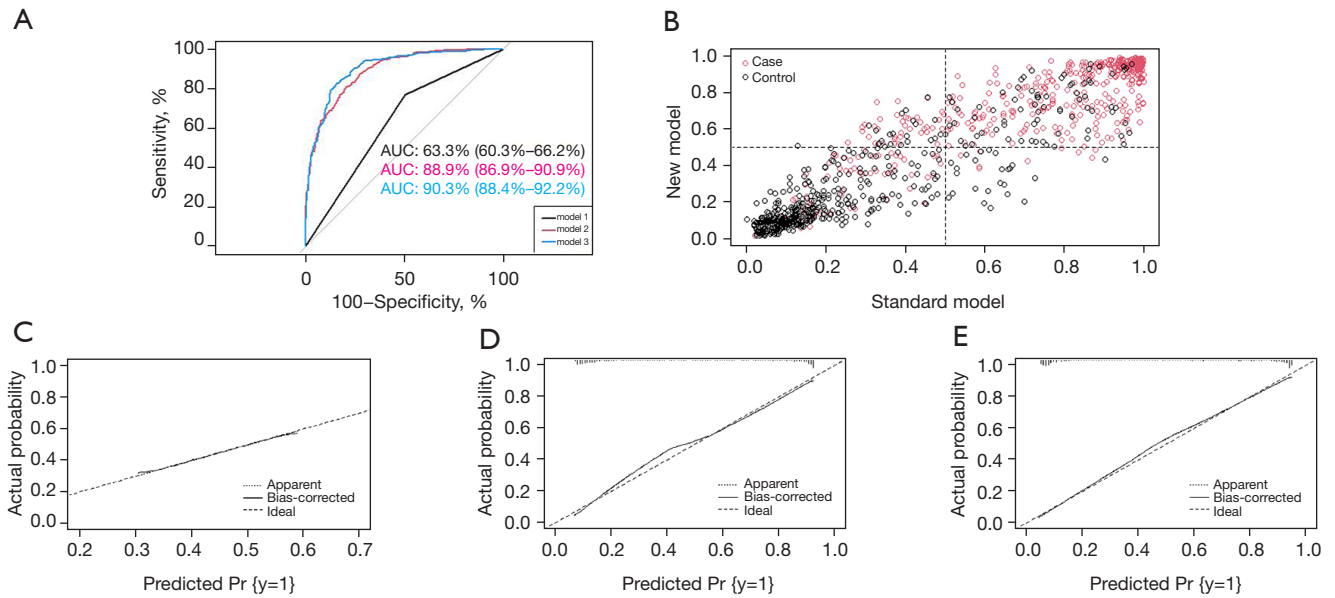


Figure 2 ROC curve, NRI, and calibration plot of the three models in the internal validation cohort. (A) AUC for the three models; (B) NRI of the three models; (C) calibration plots of Model 1; (D) calibration plots of Model 2; (E) calibration plots of Model 3. AUC, area under the ROC curve; NRI, net reclassification index; Pr, probability; ROC, receiver operating characteristic.

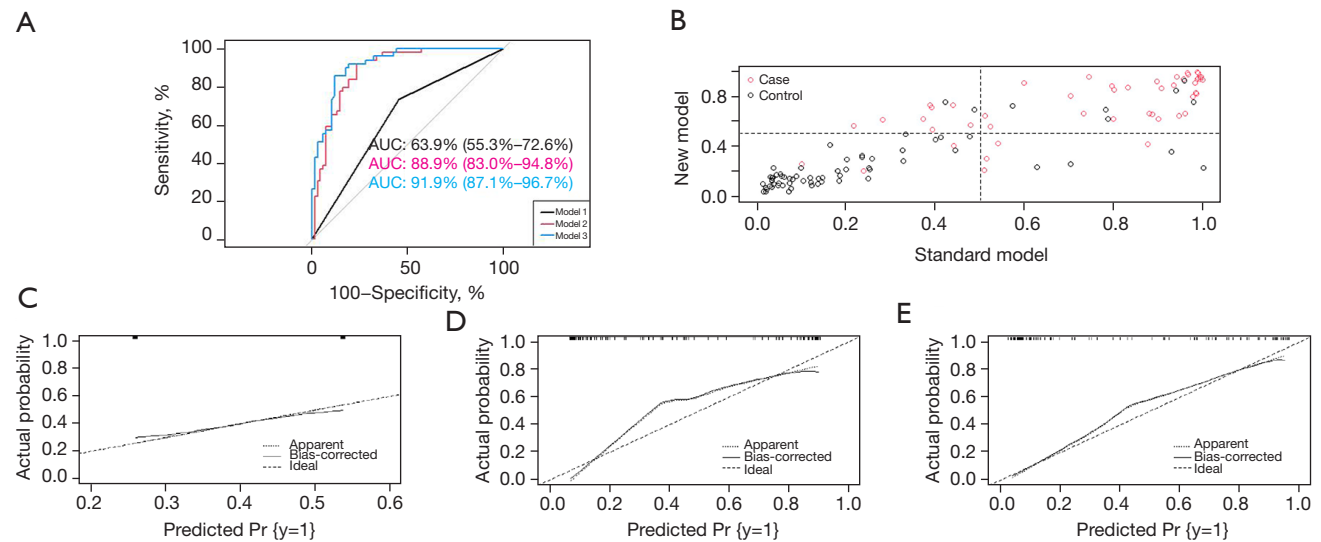


Figure 3 ROC curve, NRI, and calibration plot of the three models in the external validation cohort. (A) AUC for the three models; (B) NRI of the three models; (C) calibration plots of Model 1; (D) calibration plots of Model 2; (E) calibration plots of Model 3. AUC, area under the ROC curve; NRI, net reclassification index; Pr, probability; ROC, receiver operating characteristic.

with GGO component, those with smaller diameter, and those with MIA pathology are high risk factors for incorrect cryosection determination.

Various reports have shown that the accuracy of FS

diagnosis varies across hospitals (9,10,12,21). Large-scale medical centers, such as the Shanghai Chest Hospital (>17,000 thoracic surgeries in 2020), may have a high surgical volume, which needs critically short FS time (usually

Table 4 Diagnostic accuracy of the different models

Cohorts	Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Internal validation	Model 1	62.82	76.82	49.70	58.88	69.58
	Model 2	79.65	78.97	80.28	78.97	80.28
	Model 3	82.76	84.98	80.68	80.49	85.14
External validation	Model 1	62.39	73.47	54.41	53.73	74.00
	Model 2	82.05	77.55	85.29	79.17	84.06
	Model 3	87.18	85.71	88.24	84.00	89.55

NPV, negative predictive value; PPV, positive predictive value.

<30 min). Consequently, the accuracy of FS pathology was measured as 61.3% in this study, which slightly improved to 65.3% when tumors were stratified based on the high- or low-risk group.

In the real-world setting, the diagnosis of “atypia, defer to permanent sections” when examining minute pulmonary lesions on FS is often made by the surgical pathologist because it avoids possible diagnostic errors and potential medico-legal exposure. In this study, the rate of such cases was as high as 33.4%. While many pathologists have also tried to adopt new methods, such as the inflation method, to improve the accuracy of FS (22,23), the number of cases in these studies was too limited to establish a definitive method of FS.

Interestingly, correct surgical extent was determined in 81.2% of patients, suggesting that even with ambiguous FS results, surgeons made partly accurate judgments, either empirically or with reference to other factors. Studies have shown that combining intraoperative FS results with tumor diameter may significantly increase judgment accuracy (12), and some investigators have also used radiomics methods combined with intraoperative FS to determine the infiltration degree of adenocarcinoma (24,25).

ML has played an increasingly important role in the classification and prediction of problems and has achieved excellent results in the diagnosis and treatment of heart failure (26), survival prediction in patients with breast cancer (27), medical imaging (28), and biomedicine (29). Meanwhile, RF, an ensemble learning method based on a decision tree, has exhibited unparalleled accuracy among current algorithms, run efficiently on large databases, and generated an internal unbiased estimate of the generalization error as forest building progresses, making it an effective method for estimating missing data while maintaining accuracy (30). As the established model

was robust and could deal with nonlinear problems, we were motivated to investigate whether a more accurate determination of SPNs could be achieved using logistic regression analysis and an RF algorithm to build models.

In clinical practice, classification of tumors into low- and high-risk groups is sufficient for surgeons. Therefore, the models were also set to determine the low- and high-risk groups rather than accurate pathological results. In the selection of clinical features, cigarette smoking was identified as a major risk factor for lung cancer because cigarettes contain numerous carcinogens, mutagens, and other toxicants (31). Regarding preoperative imaging, we selected GGO component and pleural indentation as two indicators, as most of these are associated with lung malignancies (32,33). Additionally, increases in tumor marker levels have been associated with certain lung malignancies (34,35). Model 2 (logistic regression), combining clinical features and FS results, and Model 3 (established using the RF algorithm) were better than Model 1. Model 3 was optimal, showing an increase in accuracy from 62.82% to 82.76% in the internal validation cohort, with significant improvements in precision and specificity. In the ROC, the AUC also increased from 63.3% to 90.3% in the internal validation cohort, and the calibration plots and NRI confirmed these results. In the independent external validation group, Model 3 increased the accuracy from 62.39% to 87.18%, and the difference was statistically significant, and in ROC, the AUC increased from 63.9% to 91.9%, as did the calibration plots and NRI. By testing the internal validation group against the external validation group, we found that Model 3 presents a significant advantage in determining the low-/high-risk group.

Therefore, we conclude that single-dimensional information (such as FS, CT, and others) is insufficient

to determine the nature of SPN more accurately, and the combination of multi-dimensional data is required to make a synergistic judgment and improve accuracy. Furthermore, as the RF algorithm-based models in ML may significantly improve the validity of the judgment, this method may effectively help surgeons decide on the surgical resection area under the current situation in which imaging histology and CT image texture analysis are not widely used.

Study limitations

First, the judgment accuracy of Model 3 was insufficient at 82.7%, although this may be improved by increasing the amount of data when using the RF algorithm. Second, the imaging features analyzed in this study included only the GGO components and pleural indentations because these features are easily accessible in real-world clinical practice and helpful in both large medical centers and small hospitals. However, advancements in radiomics techniques may allow the use of vast information contained within CT images in future studies. Notably, deep learning techniques offer a potential solution for interpreting these complex and ever-increasing data in CT images. Our previous study identified epidermal growth factor receptor mutation status in patients with lung adenocarcinoma using CT images based on a three-dimensional deep convolutional neural network method (36). The application of deep learning and extraction of additional CT data may improve model accuracy in this study. Third, we validated the classification results of the model using an external validation cohort, and the results showed Model 3 still exhibited the best classification results, and a larger AUC was obtained compared with Model 3 in the internal validation cohort. However, the AUC in Model 3 was not significantly larger than that in Model 2 because of the relatively small number of cases included in the external validation cohort. Therefore, it is reasonable to suspect Model 3 may exhibit better classification results when applied to a larger external population. Future studies may overcome these limitations by conducting multicenter, standardized trials and exploring more suitable ways of combining large amounts of clinical data and FS to identify strategies that may increase the accuracy of intraoperative classification in patients with SPNs.

In conclusion, our results suggest an RF model combining clinical characteristics and intraoperative FS may significantly improve the accuracy of SPN classification. The model may also be used as a reliable complementary

method when FS evaluation is equivocal, resulting in a more accurate extent of surgical resection. This may aid surgeons in making more accurate surgical decisions to avoid unnecessary lung function loss and related complications. Future studies should consider using deep learning to quantitatively analyze paraffin sections (used to determine neurological tumor pathology) (37) and intraoperative FS images to incorporate them into the model, improving the model accuracy and increasing the objectivity of intraoperative FS analysis.

Acknowledgments

The authors appreciate the academic support from the AME Thoracic Surgery Collaborative Group. We wish to thank Editage (www.editage.cn) for English language editing.

Funding: This work was supported by the Major Research Plan of the National Natural Science Foundation of China (grant number 92059206). The Foundation sponsored a part of the data collection, random forest algorithm, and statistical analyses cost.

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-22-395/rc>

Data Sharing Statement: Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-22-395/dss>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-22-395/coif>). XF reports funding from the National Natural Science Foundation of China (grant number 92059206). The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study design is approved by the Committee of Medical Ethics of Shanghai Chest Hospital (approval number KS21002, 2021-2) and Hwa Mei Hospital, University of Chinese Academy of Science

(Ningbo No. 2 Hospital, approval number YJ-NBEY-KY-2021-140-01, 2021-9). Informed consent was waived because of the retrospective nature of the study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Cruickshank A, Stieler G, Ameer F. Evaluation of the solitary pulmonary nodule. *Intern Med J* 2019;49:306-15.
2. Oda S, Awai K, Murao K, et al. Volume-doubling time of pulmonary nodules with ground glass opacity at multidetector CT: Assessment with computer-aided three-dimensional volumetry. *Acad Radiol* 2011;18:63-9.
3. Ginsberg RJ, Rubinstein LV. Randomized trial of lobectomy versus limited resection for T1 N0 non-small cell lung cancer. Lung Cancer Study Group. *Ann Thorac Surg* 1995;60:615-22; discussion 622-3.
4. Travis WD, Brambilla E, Noguchi M, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* 2011;6:244-85.
5. Zhang Y, Ma X, Shen X, et al. Surgery for pre- and minimally invasive lung adenocarcinoma. *J Thorac Cardiovasc Surg* 2022;163:456-64.
6. Caplin ME, Baudin E, Ferolla P, et al. Pulmonary neuroendocrine (carcinoid) tumors: European Neuroendocrine Tumor Society expert consensus and recommendations for best practice for typical and atypical pulmonary carcinoids. *Ann Oncol* 2015;26:1604-20.
7. Ye T, Deng L, Xiang J, et al. Predictors of Pathologic Tumor Invasion and Prognosis for Ground Glass Opacity Featured Lung Adenocarcinoma. *Ann Thorac Surg* 2018;106:1682-90.
8. Mazzone P, Jain P, Arroliga AC, et al. Bronchoscopy and needle biopsy techniques for diagnosis and staging of lung cancer. *Clin Chest Med* 2002;23:137-58, ix.
9. Liu S, Wang R, Zhang Y, et al. Precise Diagnosis of Intraoperative Frozen Section Is an Effective Method to Guide Resection Strategy for Peripheral Small-Sized Lung Adenocarcinoma. *J Clin Oncol* 2016;34:307-13.
10. Walts AE, Marchevsky AM. Root cause analysis of problems in the frozen section diagnosis of in situ, minimally invasive, and invasive adenocarcinoma of the lung. *Arch Pathol Lab Med* 2012;136:1515-21.
11. Yeh YC, Nitadori J, Kadota K, et al. Using frozen section to identify histological patterns in stage I lung adenocarcinoma of ≤ 3 cm: accuracy and interobserver agreement. *Histopathology* 2015;66:922-38.
12. Zhu E, Xie H, Dai C, et al. Intraoperatively measured tumor size and frozen section results should be considered jointly to predict the final pathology for lung adenocarcinoma. *Mod Pathol* 2018;31:1391-9.
13. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25:30-6.
14. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* 2018;33:2358-61.
15. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35:1925-31.
16. Henschke CI, Yankelevitz DF, Altorki NK. The role of CT screening for lung cancer. *Thorac Surg Clin* 2007;17:137-42.
17. Gu Z, Wang H, Mao T, et al. Pulmonary function changes after different extent of pulmonary resection under video-assisted thoracic surgery. *J Thorac Dis* 2018;10:2331-7.
18. Okada M, Koike T, Higashiyama M, et al. Radical sublobar resection for small-sized non-small cell lung cancer: a multicenter study. *J Thorac Cardiovasc Surg* 2006;132:769-75.
19. Eguchi T, Kameda K, Lu S, et al. Lobectomy Is Associated with Better Outcomes than Sublobar Resection in Spread through Air Spaces (STAS)-Positive T1 Lung Adenocarcinoma: A Propensity Score-Matched Analysis. *J Thorac Oncol* 2019;14:87-98.
20. Miller DL, Rowland CM, Deschamps C, et al. Surgical treatment of non-small cell lung cancer 1 cm or less in diameter. *Ann Thorac Surg* 2002;73:1545-50; discussion 1550-1.
21. Marchevsky AM, Changsri C, Gupta I, et al. Frozen section diagnoses of small pulmonary nodules: accuracy and clinical implications. *Ann Thorac Surg* 2004;78:1755-9.
22. Xu X, Chung JH, Jheon S, et al. The accuracy of frozen section diagnosis of pulmonary nodules: evaluation of inflation method during intraoperative pathology consultation

- with cryosection. *J Thorac Oncol* 2010;5:39-44.
23. Xiang Z, Zhang J, Zhao J, et al. An effective inflation treatment for frozen section diagnosis of small-sized lesions of the lung. *J Thorac Dis* 2020;12:1488-95.
 24. Wu G, Woodruff HC, Sanduleanu S, et al. Preoperative CT-based radiomics combined with intraoperative frozen section is predictive of invasive adenocarcinoma in pulmonary nodules: a multicenter study. *Eur Radiol* 2020;30:2680-91.
 25. Wang B, Tang Y, Chen Y, et al. Joint use of the radiomics method and frozen sections should be considered in the prediction of the final classification of peripheral lung adenocarcinoma manifesting as ground-glass nodules. *Lung Cancer* 2020;139:103-10.
 26. Olsen CR, Mentz RJ, Anstrom KJ, et al. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *Am Heart J* 2020;229:1-17.
 27. Kalafi EY, Nor NAM, Taib NA, et al. Machine Learning and Deep Learning Approaches in Breast Cancer Survival Prediction Using Clinical Data. *Folia Biol (Praha)* 2019;65:212-20.
 28. Erickson BJ, Korfiatis P, Akkus Z, et al. Machine Learning for Medical Imaging. *Radiographics* 2017;37:505-15.
 29. Goecks J, Jalili V, Heiser LM, et al. How Machine Learning Will Transform Biomedicine. *Cell* 2020;181:92-101.
 30. Yang L, Wu H, Jin X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. *Sci Rep* 2020;10:5245.
 31. Song MA, Benowitz NL, Berman M, et al. Cigarette Filter Ventilation and its Relationship to Increasing Rates of Lung Adenocarcinoma. *J Natl Cancer Inst* 2017.
 32. Kobayashi Y, Mitsudomi T. Management of ground-glass opacities: should all pulmonary lesions with ground-glass opacity be surgically resected? *Transl Lung Cancer Res* 2013;2:354-63.
 33. Kim HJ, Cho JY, Lee YJ, et al. Clinical Significance of Pleural Attachment and Indentation of Subsolid Nodule Lung Cancer. *Cancer Res Treat* 2019;51:1540-8.
 34. Grunnet M, Sorensen JB. Carcinoembryonic antigen (CEA) as tumor marker in lung cancer. *Lung Cancer* 2012;76:138-43.
 35. Karnak D, Ulubay G, Kayacan O, et al. Evaluation of Cyfra 21-1: a potential tumor marker for non-small cell lung carcinomas. *Lung* 2001;179:57-65.
 36. Xiong JF, Jia TY, Li XY, et al. Identifying epidermal growth factor receptor mutation status in patients with lung adenocarcinoma by three-dimensional convolutional neural networks. *Br J Radiol* 2018;91:20180334.
 37. Khalsa SSS, Hollon TC, Adapa A, et al. Automated histologic diagnosis of CNS tumors with machine learning. *CNS Oncol* 2020;9:CNS56.
- (English Language Editor: B. Draper)

Cite this article as: Qian L, Zhou Y, Zeng W, Chen X, Ding Z, Shen Y, Qian Y, Tosi D, Silva M, Han Y, Fu X. A random forest algorithm predicting model combining intraoperative frozen section analysis and clinical features guides surgical strategy for peripheral solitary pulmonary nodules. *Transl Lung Cancer Res* 2022;11(6):1132-1144. doi: 10.21037/tlcr-22-395