Development and Validation of an Explainable Machine Learning-Based Model for Predicting the Interval Growth of Pulmonary Subsolid Nodules: A Prospective Multicenter Cohort Study

Zhedong Zhang[1,2,3#], Chao Sun[1,4#], Lixin Zhou[1,2], Hao Li[1,2], Xianjun Min[5], Jiabao Liu[6], Zhitao Wang[7], Dawei Wang[8], Nan Hong[1,4*], Fan Yang[1,2*], Xiao Li[1,2*]

[1] Department of Thoracic Surgery, Peking University People's Hospital, Beijing 100044, PR China.

[2] Thoracic Oncology Institute, Peking University People's Hospital, Beijing 100044, PR China.

[3] Department of Thoracic Surgery, Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang University, Hangzhou 310009, PR China.

[4] Department of Radiology, Peking University People's Hospital, Shijiazhuang 050051, PR China.

[5] Department of Thoracic Surgery, AMHT Group Aerospace 731 Hospital, Beijing 100044, PR China.

[6] Department of Thoracic Surgery, Shijiazhuang People's Hospital, Beijing 100044, PR China.

[7] Yijian Dian Health Management Center, Peking University Health Science Center, Beijing 100044, PR China.

[8] Institute of Advanced Research, Infervision Medical Technology Co., Ltd, Beijing 100025, PR China.

Zhedong Zhang and Chao Sun contributed equally to this work and shared the first authorship.

* Corresponding authors at: Peking University People's Hospital, No. 11 Xizhimen South Street, Xicheng District, Beijing 100044, China (Nan. H). Peking University People's Hospital, No. 11 Xizhimen South Street, Xicheng District, Beijing 100044,

China (F. Yang). Peking University People's Hospital, No. 11 Xizhimen South Street, Xicheng District, Beijing 100044, China (X. Li).
E-mail address: dr.lixiao@163.com (X. Li)

Abstract

Objectives: In this multicenter study, we aimed to develop and validate a predictive model for pulmonary subsolid nodules (SSN) growth at different time intervals by machine learning (ML) based CT radiomics methods. This model is intended to guide personalized follow-up strategies in clinical practice.

Methods: A total of 642 patients with 717 SSNs who underwent long-term follow-up were retrospectively collected from three medical centers. Patients were categorized into growth and non-growth groups based on the growth status of presented SSNs within 2 or 5 years, and they were randomly divided into training and internal testing sets at an 8:2 ratio. Predictive models were developed using the optimal ML algorithms for clinical, radiomics, and clinical-radiomics fusion models to assess the risk of SSN growth over different timeframes. An independent external test set was established by including another 95 patients with 105 SSNs from a health examination center. Multiple assessment indices, including the area under the receiver-operating-characteristic curve (AUC), were utilized to assess and compare predictive performance. Furthermore, the SHapley Additive exPlanation (SHAP) method was employed to rank the importance of features and elucidate the rationale behind the final model.

Results: The extreme gradient boosting (XGBoost) and light gradient boosting machine (Light GBM) model performed best in discriminative ability among the 8 ML models. For the prediction of within-2-year growth, the clinical, radiomics, and clinical-radiomics fusion models developed using the optimal ML algorithms achieved the AUC of 0.823 (95% CI: 0.745-0.906), 0.889 (95% CI: 0.823-0.943), and 0.911 (95% CI: 0.858-0.955) on the internal testing set, and the AUC of 0.712 (95% CI: 0.610-0.815), 0.734 (95% CI: 0.616-0.830), and 0.734 (95% CI: 0.623-0.835) on the external testing set. In 5-year growth prediction task, the three models achieved AUCs of 0.796 (95% CI: 0.708-0.884), 0.838 (95% CI: 0.759-0.905), and 0.849 (95% CI: 0.772-0.913) on the internal testing set and AUCs of 0.672 (95% CI: 0.550-0.795), 0.773 (95% CI: 0.657-0.880), and 0.776 (95%

CI: 0.652-0.882) on the external testing set. Furthermore, these insights have been translated into a streamlined clinical management framework, enhancing its utility within clinical settings.

Conclusions: The interpretable machine learning model we developed based on multicenter longitudinal follow-up data for SSN has been successfully developed to accurately predict changes in SSN over 2 years and used for the first time to guide 5-year long-term follow-up.

Keywords: Subsolid nodules, Natural course, Lung adenocarcinoma, Radiomics, Machine learning.

1. Introduction

Lung cancer is the most common cancer globally, and with the widespread use of Low-Dose Computed Tomography (LDCT), the detection rate of pulmonary subsolid nodules (SSN) has significantly increased[1]. Previous research has indicated that the appearance of growth during follow-up strongly suggests malignancy in SSN[2, 3]. Recent studies have shown that SSNs that persist and remain stable for over five years still require continued monitoring, and extended follow-up beyond five years may reveal more cases of lung cancer[4-7].

Due to differences in the biological characteristics and prognosis of pure ground-glass nodules (pGGN) and part-solid nodules (PSN), major guidelines manage these two types separately. However, there is considerable controversy in determining the presence of solid components within SSN and measuring the solid components[8]. Some PSNs have solid components visible in both the lung and mediastinal windows (real part-solid nodules, rPSN). In contrast, others have solid components only visible in the lung window (heterogeneous ground-glass nodules, hGGN). Previous research has reported that the average time for hGGN to develop into rPSN is 2.1 years[9], and rPSN has poorer clinical pathological results and prognosis than hGGN[10]. Previous studies on the natural course of SSN have primarily focused on comparing differences between pGGN and PSN based on lung window classification[11-13]. However, these studies have not provided sufficient information on the natural course of solid components under different window settings. Current guidelines for SSN management have several limitations. Therefore, establishing personalized SSN management methods and developing appropriate follow-up strategies hold significant clinical importance.

Radiomics, as a non-invasive method, can extract numerous features from Computed Tomography (CT) scan images through high-throughput computation, transform them into

comprehensive quantifiable data, and develop models to non-invasively predict various phenotypic features of lesions[14, 15]. Numerous radiomics studies have established predictive models to enhance the accuracy of diagnosing benign and malignant nodules, assessing the degree of infiltration, and predicting histological subtypes and prognosis in lung cancer patients[16]. However, due to limitations in the number of cases and follow-up time, how to dynamically track nodules and predict the growth patterns of SSNs using radiomics remains to be further explored [17, 18].

Based on these scientific questions, this study selected SSN patients from multiple centers with long-term follow-up as the research subjects, including those with growth at two years and five years of follow-up and those with sustained stability. Through ML modeling, clinical and radiomics features of patients were used to establish clinical-radiomics fusion models to predict the growth status of SSNs within two years and five years, optimizing individualized follow-up strategies.

2. Material and Methods

2.1 Study Design and Inclusion Criteria

This study is a multi-center retrospective cohort study based on the STORBE guidelines. It includes adult individuals aged 18 years or older who underwent chest CT scans for any reason at three tertiary comprehensive medical centers from November 2007 to August 2021, regardless of their smoking history. They were randomly divided into a training set and an internal testing set in an 8:2 ratio. In addition, patients from a medical examination center in Beijing, China, who underwent chest CT examinations between January 2017 and November 2021 were included as an independent external testing set. The inclusion and exclusion criteria were identical to those of the derivation cohort. **(Detailed information on participating hospitals was listed in Supplementary Materials)**. Clinical data, such as smoking history, previous history of pulmonary conditions and malignancies, and chest CT scans of enrolled patients, were all retrospectively collected for modeling. Notably, our study had two key components **(Figure 1)**. First, we constructed predictive models for growth at different time intervals. We pretested the probability of nodule growth using eight typical machine learning algorithms and selected the best-performing algorithm for further optimization. Secondly, the model underwent external validation in an independent external SSN cohort to assess its performance.

The inclusion criteria for SSN were as follows: (1) confirmation of SSN persisting for at least

six months after the initial chest CT examination; (2) SSN with a maximum long-axis diameter of ≤3 cm; (3) evaluation of each SSN using original Digital Imaging and Communications in Medicine (DICOM) format files from chest CT with a slice thickness of ≤1.5 mm; (4) a minimum follow-up time of at least two years or a follow-up time of less than two years but with documented nodule growth during the follow-up period. Exclusion criteria were: (1) inability to obtain detailed clinical data for patients with SSN; (2) a decrease in the maximum long-axis diameter of the SSN by ≥2 mm during the follow-up period.

Non-smokers were defined as individuals who had never smoked in their lifetime or had smoked fewer than 100 cigarettes. Individuals with data on smoking status but without data on the quantity of cigarettes smoked were included. Exclusion criteria included participants with a history of lung cancer at the baseline screening and those with unknown smoking status history.

This study obtained approval from the Institutional Review Boards (IRB) of three tertiary comprehensive medical centers (No. 2022PHB031-001, No. 2022-0601-01, and No. 2022002). The IRB waived the requirement for written informed consent from the participants.

2.2 Growth Definition and SSNs Categorization

SSN growth is defined as occurring during follow-up[19]: (1) an increase in the maximum diameter by ≥2mm; (2) an increase in the solid component of PSN (including hGGN and rPSN) by ≥2mm; (3) the appearance of any diameter of a solid component in pGGN under the lung window. SSN reduction is a decrease in the maximum diameter of the solid component by ≥2mm during follow-up[20]. SSN stability is defined as not meeting the criteria for growth or reduction during follow-up. SSN is categorized into pGGN, hGGN, and rPSN based on the radiological features of lung and mediastinal windows in chest CT.

Given that the 2-year and 5-year progression-free survival serve as important prognosis indicators for patients with cancer[4, 5, 12], we accordingly studied the growth risks of SSN within 2-year and 5-year periods. In particular, the enrolled patients were categorized according to the status of SSN (growth or stable) within two years and five years, respectively. SSNs were labeled as positive if they grew within the 2 or 5-year follow-up period and assigned as negative if they remained stable during the 2 or 5-year follow-up period, regardless of subsequent growth afterward.

2.3 Clinical and Radiographic Data Collection

We collected primary clinical, surgical, and pathological data of SSN patients included in the study. This had demographic information, surgical procedures, and the pathological diagnosis of SSNs in patients who underwent resection surgery. A chief thoracic surgeon or higher decided for the need for invasive tests to obtain pathological results for SSNs. The pathological types of SSNs were categorized as benign, lung adenocarcinoma (LUAD) (invasive adenocarcinoma, IAC, and minimally invasive adenocarcinoma, MIA), and their epithelial precursor lesions (adenocarcinoma in situ, AIS and atypical adenomatous hyperplasia, AAH).

We gathered imaging data for the included SSNs, encompassing both CT quantitative features and non-quantitative features. The detailed information regarding CT acquisition and parameters is described in **Figure S1**. This data had the maximum diameter of SSNs, the diameter of the solid component under the lung window (LW) and mediastinal window (MW), the Lung Window-Consolidation Tumor Ratio (LW-CTR), the Mediastinal Window-Consolidation Tumor Ratio (MW-CTR), and the number of multifocal SSNs. We also collected information on SSN types, lobulation signs, air bronchograms, vascular signs, pleural tag signs, and others. In discrepancies, a consensus was reached through discussion with a third radiologist.

2.4 Nodules Segmentation and Feature Selection

A radiology-trained thoracic surgeon (Reader 1, with ten years of chest imaging experience) and a radiologist (Reader 2, with 15 years of experience in reading chest imaging) independently performed layer-by-layer SSN segmentation based on radiological examination reports. The ROI of each nodule layer was manually drawn based on the image using 3D Slicer (version 5.2.1) **(Figure S2)**. If there were multiple SSNs, the largest one was selected for analysis. To ensure the stability of feature extraction, the consistency of radiomic features was assessed between different observers and by the same observer at other times. Three months later, 30 nodules were randomly selected and segmented again by the above two readers. Two readers were blinded to clinical characteristics, SSN growth, and histopathology.

A voxel size standardization of 1mm was applied along the x, y, and z axes. Pyradiomics (version 3.0.1) was used to extract SSN radiomic features, resulting in a total of 1454 radiomic features, including first-order features, Gray Level Cooccurrence Matrix (GLCM) features, Gray

Level Dependence Matrix (GLDM) features, Gray Level Run Length Matrix (GLRLM) features, Gray Level Size Zone Matrix (GLSZM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features, and shape-based features. Detailed information on extracted features was summarized in **Supplementary Table 1**.

Clinical features, including clinical characteristics and conventional imaging features, were selected for subsequent model construction through univariate and multivariate logistic regression.

2.5 Model Construction, Selection, and Validation

Eight machine learning models, including Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (Light GBM), K-Nearest Neighbor (KNN), and Multilayer Perceptron (MLP), were trained with selected clinical and radiomic features to build three sets of models: clinical models, radiomic models, and clinical-radiomic fusion models. The models with the highest diagnostic performance parameters were selected using repeated three-fold cross-validation on the training dataset. The model with the best AUC calculated on the internal testing dataset was used as the final model for application on the external testing dataset. Discrimination was quantified using the area under the curve (AUC). Several commonly used evaluation indexes, such as the area under the receiver-operating-characteristic (ROC) curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F1 score, were used to evaluate the reliability of these models. Predictive accuracy was assessed using calibration curves and confusion matrices. Shapley Additive exPlanations (SHAP) was used to visualize the correlations between variables and SSN growth.

2.6 Follow up

The growth interval refers to the time from the baseline chest CT scan to the subsequent follow-up CT scan, during which the same SSN met the criteria for growth. The total observation time is the interval between the baseline and final CT scans for the same SSN or between the baseline CT scan and the SSN's last intervention. The follow-up CT intervals were determined by specialized clinical thoracic surgeons based on patient and SSN radiological features, following guidelines.

All patients were followed up via phone or outpatient visits, and the outcomes of the nodules were recorded. The final follow-up deadline was December 2022.

2.7 Statistical analysis

Statistical analysis was conducted using SPSS (version 26), R software (version 3.6.2), and Python (version 3.11). For data following a normal distribution, values were presented as mean ± standard deviation (SD), and intergroup comparisons were performed using independent sample t-tests. Data with non-normal distribution were described as median [interquartile range, IQR] and analyzed using the Mann-Whitney U test. Categorical variables were presented as frequencies and percentages, and intergroup comparisons were made using the Chi-square test or Fisher's exact test when appropriate. The DeLong test was employed to compare different ROC curves. Feature selection for radiomics and model construction was made using Python's "scikit-learn" machine learning framework. ROC curves and confusion matrices were generated using Python's "Matplotlib" library. A significance level of $p < 0.05$ was considered for all tests.

3. Results

3.1 Baseline Clinical Characteristics of the Patients and Radiologic Features of Included SSNs

This study included 642 patients with 717 SSNs from three different hospitals and 99 patients with 105 SSNs from one medical examination center. Two-year and five-year growth prediction models were established based on whether SSNs grew within 2 and 5 years, respectively. Clinical baseline characteristics of all patients and radiological characteristics of all SSNs are summarized in **Table 1** and **Table 2**.

All patients and SSNs from three tertiary comprehensive medical centers were randomly divided into training and internal test sets in an 8:2 ratio. Patient information and SSN features for the two datasets are compared in **Table S2 and Table S3.** There were no significant differences in the pathological characteristics and surgical methods of SSNs that underwent surgical resection in the training and internal testing sets; details are provided in **Table S4**.

3.2 Clinical Feature selection, model development, and performance comparison

The univariate and multivariate analyses revealed that gender, SSN type, vascular sign,

and initial maximum diameter were independent risk factors for the different-year prediction models. In the 5-year prediction model, the vacuole sign was a newly discovered factor **(Table S5)**. The clinical models established based on the optimal machine learning algorithm, XGBoost, achieved an AUC of 0.823 (95% CI: 0.745-0.906) and 0.796 (95% CI: 0.708-0.884) in the internal testing cohort for the 2-year and 5-year predictions, respectively (Comparison of Clinical Models Established by Different Machine Learning Algorithms and ROC Curves are presented in Supplementary Material **Figure S3** and **Table S6**).

3.3 Radiomics Feature selection, model development, and performance comparison

Among the 1,454 radiomic features, redundant features (ICC < 0.75 and PCC > 0.9) were first removed, resulting in 271 and 259 remaining features for the 2-year and 5-year models, respectively. The final set of 10 radiomic features was generated through recursive feature elimination and cross-validation **(Table S7)**. For the 2-year model, the radiomic model was developed using the optimal machine learning algorithm LightGBM, achieving an AUC of 0.889 (95% CI: 0.823-0.943) in the internal testing cohort. For the 5-year model, the radiomic model was developed using the optimal machine learning algorithm XGBoost, achieving an AUC of 0.838 (95% CI: 0.759-0.905) in the internal testing cohort **(Figure S4 and Table S8).**

3.4 Development of fusion radiomics model and performance comparison

Incorporating the selected clinical features and radiomic features, a clinical-radiomic fusion model was developed using the mentioned eight machine learning algorithms (**Figure S5 and Table S9**). The results show that for the 2-year prediction model, the LightGBM algorithm with its specific parameters performed the best on the internal testing dataset, achieving an AUC of 0.911 (95% CI: 0.858-0.955). For the 5-year prediction model, the XGBoost algorithm with its specific parameters performed the best on the internal testing dataset, resulting in an AUC of 0.849 (95% CI: 0.772-0.913).

3.5 Identification of the Final Model and External Validation

The optimal algorithm with the highest AUC in the internal testing set was selected to build the model, and its predictive performance was compared. **Table 3** displays the predictive

performance of the three groups of prediction models, and ROC curves are presented in **Figure 2**. The addition of SHAP allows for interpretative analysis of the fused radiomic model by visualizing the specific impact of each variable on the prediction of SSN growth **(Figure 3).** In the 2-year prediction model, the fusion model produced the highest AUC in the internal testing cohort. Both the fusion model and radiomics model had higher AUC values compared to the clinical model (DeLong test $p < 0.05$), with no significant difference between the fusion model and radiomics model (DeLong test $p > 0.05$). In the internal testing set of the 5-year prediction model, even though there were no statistically significant differences among the three models, the fusion model achieved the highest AUC. Simultaneously, the radiomic and fusion models exhibit higher accuracy and sensitivity.

In the external testing set of the 2-year prediction model, the AUC values for the clinical model, radiomics model, and clinical-radiomics fusion model were 0.712 (95% CI: 0.610-0.815), 0.734 (95% CI: 0.616-0.83), and 0.734 (95% CI: 0.623-0.835), respectively. In the external testing set of the 5-year prediction model, the AUC values for the three groups were 0.672 (95% CI: 0.550-0.795), 0.773 (95% CI: 0.657-0.880), and 0.776 (95% CI: 0.652-0.882), respectively **(Figure 4)**. The metrics for evaluating the reliability of these models are outlined in **Table 4**. The confusion matrices for each model are displayed in **Figures S6 and S7**. This suggests that the combined radiomics model exhibits good stability and reproducibility.

3.6 Follow-up management framework for clinical utility

Based on the research outcomes, we propose a tailored follow-up management framework for SSN, outlined in **Figure 5**. Firstly, for patients identified with SSN via thoracic thin-section CT, clinical physicians assess whether further examination, surgery, no follow-up, or regular follow-up is necessary. If regular follow-up is deemed necessary, the SSN is subjected to growth prediction models for evaluation. If the 2-year growth prediction model indicates a high risk of growth, clinical physicians may propose high-risk management recommendations after a thorough review. This involves extensive discussion with the patient to determine whether to continue follow-up with a shortened interval, undergo further examination, or proceed with surgery. The SSN is further evaluated using a 5-year growth prediction model if the growth risk is low. Suppose the 5-year growth prediction model indicates a high risk of SSN growth. In that

case, appropriate management recommendations are proposed after a thorough evaluation, possibly shortening the follow-up interval for continued monitoring and considering further examination if necessary. If the growth risk is low, low-risk management recommendations are provided, suggesting regular follow-up for the patient and possibly extending the follow-up interval appropriately.

4. Discussion

In clinical practice, determining the nature of nodules and establishing follow-up duration is crucial for long-term persist and stable SSN. In this study, based on multicenter, long-term follow-up cases, we constructed a growth prediction model using machine learning methods to predict the growth status of SSN at 2 and 5 years, aiming to assist in the standardized management of pulmonary nodules. This clinical prediction model demonstrates excellent diagnostic performance and has been validated using independent external data.

Pulmonary SSN growth modes were categorized into five patterns based on consecutive follow-up CT scans, including linear, rapidly accelerating, slow accelerating, slow, and rapid growth while pointing out that the likelihood of malignant radiological features of nodules increases with prolonged follow-up[18]. Therefore, index models based on volume doubling time (VDT) may not be suitable for evaluating every nodule encountered in clinical practice. Studies have shown that after two years or more of stability, the probability of SSN experiencing growth is only 5%[21]. Additionally, within three years, 26.9% of SSNs showed growth; among those stable for three years and followed up to five years, 6.7% demonstrated growth[22, 23]. Taking into account the aforementioned concerns, this study categorized subjects into groups based on their growth within two years and five years respectively. Utilizing predictive modeling, it aimed to assess growth risks and devise tailored follow-up strategies for individuals across different growth durations.

The progression of SSNs represents a complex and dynamic process. As anticipated, this study demonstrates that gender, larger size on initial CT imaging, and the presence of solid components are independent risk factors for SSN progression, consistent with previous research [5, 6, 9, 23, 24]. Introducing the radiographic characteristics of CT MW images into classification, SSNs are divided into pGGN, hGGN, and rPSN. It was found that hGGN and rPSN have

differences in growth patterns[9, 25]. Further analysis combining large-panel targeted sequencing confirmed at the genomic level that only hGGNs with solid components on LW are intermediate subtypes of PSNs. Meanwhile, the genomic structure of hGGNs is closer to pGGNs[26]. The assessment of nodule growth should not only depend on size changes but may also be influenced by morphological features. Vascular signs are defined as either vessel traversing through the lesion or vascular thickening and tortuosity around the lesion, indicating a higher demand for blood supply and often indicative of malignancy[27]. The relationship between vessels and SSNs can be classified into three types: Type I (intact vessels passing through or traversing the SSN without tiny branches), Type II (intact vessels passing through the SSN without tiny branches), and Type III (distorted vessels within the SSN are wider or tortuous). Type II and Type III are more likely to be associated with malignancy than Type I. Hence, the relationship between vessels and lesions might predict SSN progression[28]. Similarly, vacuole sign has also been established as predictive factors for the growth of SSNs in previous studies [24].

The most commonly used method currently is to select radiomic features through LASSO regression and then estimate each radiomic feature using logistic regression algorithms to calculate a Rad-score for predicting SSN growth[12, 17]. In this study, we applied dimensionality reduction to 1454 radiomic features using the REF method and built models using the ten optimal features associated with SSN growth in 2 and 5 years, which outperformed previous reports. We also found that the most influential feature in the 2-year radiomics model was glszm_HighGrayLevelZoneEmphasis_original, which assists in quantifying regions of high brightness in the image[29]. In the 5-year radiomics model, the most influential feature was glrlm_LongRunEmphasis_wavelet-LLH, which measures the frequency and intensity of continuous appearances of pixels with the same grayscale values in the image[30].

Previous studies have utilized LR on single-center retrospective data to build clinical-radiomics nomograms for predicting 2-year growth of uncertain pulmonary nodules, emphasizing the importance of combining clinical and radiomic features in predicting nodule growth. However, the inclusion of nodules confirmed by histology and a high proportion of malignant nodules in these studies led to an overestimation of the diagnostic performance of the model[17]. Similarly, Chen et al. established clinical-radiomics nomograms for predicting

the growth of SSN beyond two years based on single-center data, with the combined model significantly outperforming the clinical model but no significant difference between the combined model and the radiomics model. This study did not incorporate patients' clinical information into the analysis. [12]. Yang et al. developed several machine-learning models to predict whether lung nodules would grow within one year. They found that a LR model combining age and radiomic features performed the best (with an AUC of 0.87 in the training set and 0.82 in the validation set)[31]. However, the studies mentioned above all suffer from relatively small sample sizes and suboptimal model performance. With improved computational capabilities and storage space availability, machine learning algorithms can analyze more complex data and provide real-time output[32, 33]. XGBoost has recently become a popular algorithm, gaining recognition in various machine-learning competitions[34]. LightGBM, compared to XGBoost, has the advantage of faster training speed and lower memory usage[35]. Both machine learning methods outperform traditional linear models in terms of predictive accuracy.

Building upon the aforementioned studies, we have conducted exploratory analyses and established, for the first time, a clinical model to predict whether SSNs will grow within 5 years based on multi-center, long-term follow-up data. The results demonstrate that in the internal testing set, the AUC values for the clinical model, radiomic model, and clinical-radiomic fusion model were 0.796 (95% CI: 0.708-0.884), 0.838 (95% CI: 0.759-0.905), and 0.849 (95% CI: 0.772-0.913), respectively. According to the DeLong test, there was no significant difference among the three models, suggesting that the clinical and combined models have equivalent efficacy in predicting whether SSNs will grow within 5 years. However, the fusion model exhibited higher accuracy (0.730 vs. 0.680), sensitivity (0.767 vs. 0.517), F1 score (0.7773 vs. 0.660), and AUPRC (0.909 vs. 0.859) compared to the clinical model. Therefore, future investigations using prospective data from long-term follow-up can further explore the predictive value of the combined model in predicting SSN growth within 5 years.

Mainstream medical societies have issued guidelines for managing pulmonary nodules, but these guidelines differ in scope and emphasis. They primarily focus on factors such as malignancy probability thresholds, follow-up schedules based on imaging features, malignancy risk calculators, and the use of VDT [36-41]. To address the complexity of managing pulmonary

nodules, we utilized the SHAP method to provide insights into the inner workings of machine learning models. We developed personalized prediction models based on clinical and imaging data to enhance clinician acceptance and decision-making. Despite the initial success, when extending the model to health examination centers, there was a decrease in diagnostic performance. However, the model still exhibited stability and reliability. Therefore, in the clinical setting, clinicians should view these models as valuable tools but also consider individual patient factors and exercise flexibility in decision-making to ensure personalized and accurate diagnosis and treatment.

However, this study has some limitations. First, it is a multi-center retrospective observational study, and it may have selection bias and temporal bias. Clinical practitioners influence variations in follow-up intervals and duration at different medical centers. Second, the study only included an Asian population, which may have significant demographic differences compared to a Caucasian population. Therefore, clinical prediction models cannot be assessed for patients of different ethnicities, and further validation of the results is needed through international multi-center cohorts. Additionally, not all growing nodules were pathologically confirmed to be malignant. Hence, the criteria for surgical intervention after nodule progression warrant further exploration.

In summary, developing individualized management strategies for SSNs is crucial in clinical practice. Based on this study's distinctly different natural courses of SSNs, we combined clinical and radiomic features using machine learning algorithms to create predictive models for assessing whether SSNs will grow within two years. Compared to previous research, our models demonstrated improved predictive performance. Additionally, for the first time, we used multi-center, long-term follow-up data to establish a predictive model for SSN growth within five years, guiding long-term follow-up of SSNs. These predictive models were further validated in an external dataset, demonstrating good generalizability.

CRediT authorship contribution statement

Declaration of Competing Interest

The authors declare that no commercial or financial relationships that could be construed as

a potential conflict of interest existed during the research.

Acknowledgments

No acknowledgments.

Funding

References

[1] Aberle D R, Adams A M, Berg C D, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening [J]. The New England journal of medicine, 2011, 365(5): 395-409.

[2] Henschke C I, Yip R, Smith J P, et al. CT Screening for Lung Cancer: Part-Solid Nodules in Baseline and Annual Repeat Rounds [J]. AJR American journal of roentgenology, 2016, 207(6): 1176-84.

[3] Kakinuma R, Muramatsu Y, Kusumoto M, et al. Solitary Pure Ground-Glass Nodules 5 mm or Smaller: Frequency of Growth [J]. Radiology, 2015, 276(3): 873-82.

[4] Lee J H, Lim W H, Hong J H, et al. Growth and Clinical Impact of 6-mm or Larger Subsolid Nodules after 5 Years of Stability at Chest CT [J]. Radiology, 2020, 295(2): 448-55.

[5] Lee H W, Jin K N, Lee J K, et al. Long-Term Follow-Up of Ground-Glass Nodules After 5 Years of Stability [J]. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 2019, 14(8): 1370-7.

[6] Tang E K, Chen C S, Wu C C, et al. Natural History of Persistent Pulmonary Subsolid Nodules: Long-Term Observation of Different Interval Growth [J]. Heart, lung & circulation, 2019, 28(11): 1747-54.

[7] Herskovitz E, Solomides C, Barta J, et al. Detection of lung carcinoma arising from ground glass opacities (GGO) after 5 years - A retrospective review [J]. Respiratory medicine, 2022, 196: 106803.

[8] van Riel S J, Sánchez C I, Bankier A A, et al. Observer Variability for Classification of Pulmonary Nodules on Low-Dose CT Images and Its Effect on Nodule Management [J]. Radiology, 2015, 277(3): 863-71.

[9] Kakinuma R, Noguchi M, Ashizawa K, et al. Natural History of Pulmonary Subsolid Nodules: A Prospective Multicenter Study [J]. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 2016, 11(7): 1012-28.

[10] Lai J, Li Q, Fu F, et al. Subsolid Lung Adenocarcinomas: Radiological, Clinical and Pathological Features and Outcomes [J]. Semin Thorac Cardiovasc Surg, 2022, 34(2): 702-10.

[11] Kim Y W, Kwon B S, Lim S Y, et al. Lung cancer probability and clinical outcomes of baseline and new subsolid nodules detected on low-dose CT screening [J]. Thorax, 2021, 76(10): 980-8.

[12] Gao C, Yan J, Luo Y, et al. The Growth Trend Predictions in Pulmonary Ground Glass Nodules Based on Radiomic CT Features [J]. Frontiers in oncology, 2020, 10: 580809.

[13] Guo X, Jia X, Zhang D, et al. Indeterminate pulmonary subsolid nodules in patients with no history of cancer: growing prediction, CT pattern, and pathological diagnosis [J]. Diagnostic and interventional radiology (Ankara, Turkey), 2022, 28(3): 230-8.

[14] Chen B, Zhang R, Gan Y, et al. Development and clinical application of radiomics in lung cancer [J]. Radiation oncology (London, England), 2017, 12(1): 154.
[15] Nie K, Shi L, Chen Q, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI [J]. Clin Cancer Res, 2016, 22(21): 5256-64.
[16] Wu Y J, Wu F Z, Yang S C, et al. Radiomics in Early Lung Cancer Diagnosis: From Diagnosis to Clinical Decision Support and Education [J]. Diagnostics (Basel), 2022, 12(5).
[17] Xue L M, Li Y, Zhang Y, et al. A predictive nomogram for two-year growth of CT-indeterminate small pulmonary nodules [J]. European radiology, 2022, 32(4): 2672-82.
[18] Du Q, Peng J, Wang X, et al. Dynamic Observation of Lung Nodules on Chest CT Before Diagnosis of Early Lung Cancer [J]. Frontiers in oncology, 2022, 12: 713881.
[19] Matsuguma H, Mori K, Nakahara R, et al. Characteristics of subsolid pulmonary nodules showing growth during follow-up with CT scanning [J]. Chest, 2013, 143(2): 436-43.
[20] Hiramatsu M, Inagaki T, Inagaki T, et al. Pulmonary ground-glass opacity (GGO) lesions-large size and a history of lung cancer are risk factors for growth [J]. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 2008, 3(11): 1245-50.
[21] Wu L, Gao C, Kong N, et al. The long-term course of subsolid nodules and predictors of interval growth on chest CT: a systematic review and meta-analysis [J]. European radiology, 2022.
[22] Kobayashi Y, Fukui T, Ito S, et al. How long should small lung lesions of ground-glass opacity be followed? [J]. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 2013, 8(3): 309-14.
[23] Cho J, Kim E S, Kim S J, et al. Long-Term Follow-up of Small Pulmonary Ground-Glass Nodules Stable for 3 Years: Implications of the Proper Follow-up Period and Risk Factors for Subsequent Growth [J]. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer, 2016, 11(9): 1453-9.
[24] Qi L-L, Wu B-T, Tang W, et al. Long-term follow-up of persistent pulmonary pure ground-glass nodules with deep learning-assisted nodule segmentation [J]. European radiology, 2020, 30(2): 744-55.
[25] Zhang Z, Zhou L, Min X, et al. Long-term follow-up of persistent pulmonary subsolid nodules: Natural course of pure, heterogeneous, and real part-solid ground-glass nodules [J]. Thoracic cancer, 2023, 14(12): 1059-70.
[26] Li H, Sun Z, Xiao R, et al. Stepwise evolutionary genomics of early-stage lung adenocarcinoma manifesting as pure, heterogeneous and part-solid ground-glass nodules [J]. Br J Cancer, 2022, 127(4): 747-56.
[27] Jin C, Cao J, Cai Y, et al. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules [J]. The Journal of thoracic and cardiovascular surgery, 2017, 153(2).
[28] Qiu T, Ru X, Yin K, et al. Two nomograms based on CT features to predict tumor invasiveness of pulmonary adenocarcinoma and growth in pure GGN: a retrospective analysis [J]. Jpn J Radiol, 2020, 38(8): 761-70.
[29] Mapelli P, Bezzi C, Palumbo D, et al. 68Ga-DOTATOC PET/MR imaging and radiomic parameters in predicting histopathological prognostic factors in patients with pancreatic

neuroendocrine well-differentiated tumours [J]. Eur J Nucl Med Mol Imaging, 2022, 49(7): 2352-63.
[30] Zhang Z, Li X, Sun H. Development of machine learning models integrating PET/CT radiomic and immunohistochemical pathomic features for treatment strategy choice of cervical cancer with negative pelvic lymph node by mediating COX-2 expression [J]. Front Physiol, 2022, 13: 994304.
[31] Yang R, Hui D, Li X, et al. Prediction of single pulmonary nodule growth by CT radiomics and clinical features - a one-year follow-up study [J]. Frontiers in oncology, 2022, 12: 1034817.
[32] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine [J]. The New England journal of medicine, 2019, 380(14): 1347-58.
[33] Kourou K, Exarchos T P, Exarchos K P, et al. Machine learning applications in cancer prognosis and prediction [J]. Comput Struct Biotechnol J, 2015, 13.
[34] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California: Association for Computing Machinery; 2016. p.785–94. [J].
[35] Song Y, Zhou G, Zhou Y, et al. Artificial intelligence CT radiomics to predict early recurrence of intrahepatic cholangiocarcinoma: a multicenter study [J]. Hepatol Int, 2023.
[36] National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology. Lung Cancer Screening, version 1.2022. [Accessed October 26, 2021]. https:. Available: www.nccn.org/professionals/physician_gls/pdf/lung_screening/pdf. [J].
[37] MacMahon H, Naidich D P, Goo J M, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017 [J]. Radiology, 2017, 284(1): 228-43.
[38] National Comprehensive Cancer Network (2022) National Comprehensive Cancer Network Clinical Practice Guidelines in Non-Small Cell Lung Cancer. Version 1. https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1450. Accessed 25 Dec 2021. [J].
[39] Baldwin D R, Callister M E. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules [J]. Thorax, 2015, 70(8): 794-8.
[40] Gould M K, Donington J, Lynch W R, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines [J]. Chest, 2013, 143(5 Suppl): e93S-e120S.
[41] American College of Radiology. Lung-RADS Version 1.1. 2019. https://www.acr.org/-/media/ACR/Files/RADS/Lung-RADS/LungRADSAssessmentCategoriesv1-1.pdf?la=en. Accessed June 21, 2019 [J].