A CT-based deep learning nomogram prediction model for identification of subsolid pulmonary nodules: Distinguishing of minimally invasive adenocarcinoma from invasive adenocarcinoma

Xiangmeng Chen, Bao Feng, Yehang Chen, Xiaobei Duan, Kunfeng Liu, Kunwei Li, Chaotong Zhang, Xueguo Liu, Wansheng Long

PII: S0720-048X(21)00522-2
DOI: https://doi.org/10.1016/j.ejrad.2021.110041
Reference: EURR 110041

To appear in: European Journal of Radiology

Received Date: 23 April 2021
Revised Date: 21 September 2021
Accepted Date: 29 September 2021


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.
A CT-based deep learning nomogram prediction model for identification of subsolid pulmonary nodules: Distinguishing of minimally invasive adenocarcinoma from invasive adenocarcinoma

**Type of Article:** Original Research

Xiangmeng Chen¹#, Bao Feng²#, Yehang Chen², Xiaobei Duan³, Kunfeng Liu⁴, Kunwei Li⁴, Chaotong Zhang¹, Xueguo Liu⁺⁺, Wansheng Long¹

**Author information:**

1. Department of Radiology, Jiangmen Central Hospital, Jiangmen, Guangdong Province, 529030, PR China.
2. School of electronic information and automation, Guilin University of Aerospace Technology, Guilin City, Guangxi Province, 541004, China.
3. Department of Nuclear Medicine, Jiangmen Central Hospital, Jiangmen, Guangdong Province, 529030, PR China.
4. Department of Radiology, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, Guangdong Province, 519000, PR China.

# indicates equal contributions.

**Corresponding Author 1:**
Wansheng Long, MD
Department of Radiology, Jiangmen Central Hospital, 23#, North Road, Pengjiang Zone, Jiangmen, Guangdong Province, 529030, PR China
Work phone: 0086-0750-3165528
E-mail: jmlws2@163.com

**Corresponding Author 2:**
Xueguo Liu, MD
Department of Radiology, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, Guangdong Province, 519000, PR China.
Work phone: 0086-0750-3165528
E-mail: liuxueg@mail.sysu.edu.cn

**Other authors’ Email-addresses are as follows:**
Xiangmeng Chen (181970902@qq.com);
Bao Feng (fengbao1986.love@163.com);
Yehang Chen (1535601070@qq.com);
Xiaobei Duan (258573168@qq.com);
Kunfeng Liu (744547490@qq.com)
ABSTRACT.

Objective.

To develop and validate a deep learning nomogram (DLN) model constructed from non-contrast computed tomography (CT) images for discriminating minimally invasive adenocarcinoma (MIA) from invasive adenocarcinoma (IAC) in patients with subsolid pulmonary nodules (SSPNs).

Materials and Methods.

In total, 365 consecutive patients who presented with SSPNs and were pathologically diagnosed with MIA or IAC after surgery, were recruited from two medical institutions from 2016 to 2019. Deep learning features were selected from preoperative CT images using convolutional neural network. Deep learning signature (DLS) was developed via the least absolute shrinkage and selection operator (LASSO). New DLN integrating clinical variables, subjective CT findings, and DLS was constructed. The diagnostic efficiency and discriminative capability were analyzed using the receiver operating characteristic method and decision curve analysis (DCA).

Results.

In total, 18 deep learning features with non-zero coefficients were enrolled to develop the DLS, which was statistically different between the MIA and IAC groups. Independent predictors of DLS and lobulated sharp were used to build the DLN. The areas under the curves of the DLN were 0.889 (95% confidence interval (CI): 0.824-0.936), 0.915 (95% CI: 0.846-0.959), and 0.914 (95% CI: 0.848-0.958) in the
training, internal validation, and external validation cohorts, respectively. After stratification analysis and DCA, the DLN showed potential generalization ability.

**Conclusion.**

The DLN incorporating the DLS and subjective CT findings have strong potential to distinguish MIA from IAC in patients with SSPNs, and will facilitate the suitable treatment method selection for the management of SSPNs.

**Key Words:** deep learning, nomogram, sub-solid nodule, lung adenocarcinoma, convolutional neural network
Introduction

Subsolid pulmonary nodules (SSPNs) are nodules that are more opaque than the surrounding normal parenchyma, but are less opaque than consolidated bronchovascular structures in chest computed tomography (CT) images [1]. Persistent SSPNs pathologically correspond with primary lung invasive adenocarcinoma (IAC) and minimally invasive adenocarcinoma (MIA) [2]. The standard treatment method for IAC is still lobectomy and systematic mediastinal lymph node dissection. However, sublobar resection and lymph node biopsy are considered adequate therapy for MIA, which conserve favorable pulmonary function and improve quality of life postoperatively [3, 4]. In addition, the 5-year disease-free survival (DFS) for patients with MIA is close to 100%, which is significantly higher than the 40–85% for those with IAC [5]. Currently, it is still a challenge for a radiologist to differentiate MIA and IAC through visual evaluation of morphologic changes on CT imaging with substantial overlaps [6, 7]. Lee et al. noted that visual assessment of SSPNs based on CT features had only moderate consistency even among experienced thoracic radiologists [8]. Literature has shown that observer consistency on nodule discrimination was moderate, while on recommendations provided to patients at follow-up was relatively low [9]. This inconsistency was mainly due to the inequitable size and component on SSPNs. Therefore, it is of critical importance to develop a prediction model to efficiently discriminate MIAs from IACs, thereby guiding treatment and predicting the prognosis of patients with SSPNs.

The computer-aided three-dimensional (3D) nodule analysis of CT images has satisfactory reproducibility and computerized texture analysis is conducive to distinguishing the subtypes of pulmonary nodules [10, 11]. Radiomics based on medical images is a non-invasive approach for building appropriate models to assess
the performance of clinical diagnosis [12]. However, hand-crafted radiomics features require extensive human workload, which may attenuate the precision of tumor boundary annotation. Additionally, the inferior reproducibility and generalizability of radiomics between different CT protocols has limited its clinical application in multiple medical centers [13]. Recently, deep learning (DL), a type of machine learning, has emerged as an encouraging decision supporting approach to mechanically analyze medical images [14]. DL applies the convolutional neural network (CNN) to construct an end-to-end grouping model through studying a hierarchy of internal representations, which is distinct different from radiomics features analysis strategy [15]. Studies on DL algorithms based on thoracic CT images have been reported in various clinical settings such as the identification and reclassification of pulmonary nodules [16], the definition of malignant nodules subtypes [17, 18], and the diagnosis of diffuse lung diseases [19].

We proposed that an accurate prediction model based on DL method may be a beneficial tool for qualitative identification of SSPNs. Therefore, we constructed a CT-based DL nomogram (DLN) model to test and validate its diagnostic efficiency in discriminating MIAs from IACs.

Materials and Methods

Study Participants

The institutional review board of two medical centers approved this study, and informed consent was waived. Chest CT images were retrospectively collected from February 2016 to December 2019 in Medical Center 1. The inclusion criteria were as follows: (a) all pulmonary nodules were surgically resected and pathologically confirmed; (b) nodules on CT were presented as SSPNs and with a diameter of less
than 30 mm; (c) patients received chest CT examinations within 1 month of surgery and were available in the picture archiving and communication system; (d) thin-slice chest CT of less than 1.5 mm; (e) no history of malignant tumors; (f) no treatment history of chemotherapy or radiotherapy. In total, 247 patients were included and this study cohort was subdivided into a training cohort (n=136) and internal validation cohort (n=111) by random selection. Another group of 118 consecutive patients were selected from Medical Center 2 between February 2016 and December 2018 using the same criteria, constituting an independent external validation cohort. The overall workflow of this study is illustrated in Figure 1.

**Chest CT Protocol and CT Findings Evaluation**

Non-contrast CT examinations were performed using one of four multi-detector CT equipment: 16-scie detector CT (80 cases; Siemens Medical Solutions, Germany), Somatom Definition Force (109 cases; Siemens Medical Solutions, Germany), 64-slice AquilionOne (112 cases; Toshiba, Japan), GE Discovery CT750 HD (64 cases; GE Medical Systems, USA). All CT images were reviewed on transverse, coronal and sagittal planes by two thoracic radiologists (with 10 and 25 years of experience in thoracic radiology) who were blinded to the pathological results. Discrepancies were resolved by consensus. Detailed information is provided in Supplementary A1.

**Pathological Diagnosis**

The average interval between latest CT scanning and surgery was 13 days (range 3-28 days). Pathological diagnoses of surgical specimens were determined by two
experienced pathologists who were uninformed of the CT findings, in accordance with the 2011 edition of International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society lung adenocarcinoma classification standards [2].

DLS Construction and Validation

The DLS model based on a convolutional neural network (CNN) was built on four steps. First, in order to improve the calculation speed and accuracy and comply with data input format of CNN, a 256*256*3 square box covering the whole nodule area was obtained from the original CT images as shown in Supplementary A2 and Figure S1A.

Second, DL features refer to the output of the convolution layer and are extracted using the Keras toolkit of Python 2.7. As shown in Supplementary Figure S1B, DL features were extracted using Dense-net 121 that have 11,264 convolutional kernels [20]. To avoid overfitting, the Dense-net 121 was pre-trained with 1.3 million original images from the ImageNet dataset, using the transfer learning strategy. Then, the pre-trained model was retrained using SSPN data. The training process details are presented in Supplementary A3. Finally, the average probability of the DL features was calculated from the output of the convolution kernels in all slices. The convolution kernels came from all the convolution layers of the Dense-net 121. Therefore, a total of 11,264 DL features were extracted.

Third, each DL feature for distinguishing MIAs and IACs was assessed by the Mann-Whitney U test. To prevent overfitting, the minimum redundancy and maximum-relevance algorithm were applied to cut feature dimensions through computing feature redundancy. Finally, the top 15% features with minimum
redundancy were enrolled in the next-step analysis.

Fourth, in the training cohort, DL features with non-zero coefficients that could distinguish between MIA and IAC were identified as valuable DL features by least absolute shrinkage and selection operator (LASSO) logistic regression with 10-fold cross validation. A DLS was generated by combining the LASSO-picked features with their coefficients. Discrepancies of DLS between MIA and IAC groups were compared by the Wilcoxon-rank sum test.

In addition, a signature based on radiomics was constructed for comparison with the DLS. The details are presented in Supplementary A4.

**DLN Built and Assessment**

Variables of clinical characteristics, subjective CT findings, and DLS that reached statistical significance ($P \leq 0.05$) in univariate analyses were incorporated into multivariate regression analysis. A DLN model was formulated by combining each independent risk factor identified in multivariate analysis through a stepwise backward selection process. The DLN was calibrated and evaluated by calibration curve analysis and the Hosmer-Lemeshow test [21]. The DLN performance was evaluated by the net reclassification index (NRI) and was compared with subjective models. The area under the curve (AUC) of each prediction model was calculated in the training, internal validation, and external validation cohorts. Stratification analysis of the patient’s gender, age, CT scan equipment, slice thickness and nodule size were performed to assess the general applicability of the DLN. Finally, DCA was used to compare the DLN with other prediction models.

**Statistical Analysis**
Continuous variables such as mean patient age between the MIA and IAC groups were compared using the Wilcoxon signed-rank test. Categorical variables such as patient’s gender and radiological CT findings were compared using the Pearson’s X² test or the Fisher’s exact test, as appropriate. Inter-observer consistency of subjective CT findings were investigated by the Cohen kappa test. According to Akaike’s Information Criterion, factors with statistically significant difference (k-value > 0.60) were selected, and a subjective model was constructed by multivariate logistic regression with a stepwise backward selection of variables.

The nomogram and DCA were fulfilled by “rms” and “dca.r.”, respectively. Receiver operating characteristics (ROC) analyses were conducted for variables of statistically significant differences in multivariate analyses using “pROC.”. All statistical analyses were carried out using Python 2.7 and R3.0.1 (http://www.rproject.org). For all tests, a P value of less than 0.05 was considered statistically significant.

Results

Patient Characteristics

The clinical demographics and subjective CT findings of the study population are detailed in Table 1. There were 136, 111, and 118 cases in the training, internal validation, and external validation cohorts, respectively.

Specifically, of the 136 cases in the training cohort, 58 were MIA and 78 were IAC. No significant differences were detected in gender and nodule location between the MIA and IAC groups (P=0.122 and P=0.954, respectively). Patients with IAC were older and more likely to present lesions with part-solid, larger size, irregular margin, and lobulated sharp, spiculated sign and air-bronchogram sign (all P<0.001).
Interobserver agreement analysis for various subjective CT findings of SSPNs showed that the consistency was high with regard to lesion size and nodule type (k=0.848 and 0.837, respectively). Meanwhile, substantial agreement was achieved in lobulated sharp, nodule margin, spiculated sign, and air-bronchogram sign (k=0.745, 0.711, 0.696, and 0.653, respectively).

Multivariate analysis identified nodule type (odds ratio (OR): 4.025; 95% confidence interval (CI), 1.626-9.965; \( P = 0.003 \)), size (OR: 3.279; 95% CI, 1.187-9.061; \( P = 0.022 \)) and lobulated sharp (OR: 3.078; 95% CI, 1.234-7.677; \( P = 0.016 \)) as the independent differentiators of MIA from IAC groups. A subjective model was constructed based on these factors.

**Radiomics Signature Construction**

Of the 10,324 radiomics features, 6893 had ICC values of greater than 0.75 and showed remarkable differences between the IMA and IAC groups in the training cohort. A total of 1,034 radiomics features were further screened out using the standard of minimum redundancy maximum relevance. Finally, 12 features with non-zero coefficients were identified by LASSO analysis. These 12 radiomics features were used to construct a formula for calculating radiomics signature scores, which were significantly higher in IAC groups as compared with MIA groups in the training, internal validation, and external validation cohorts (all \( P < 0.001 \); Table 1). The predictive efficiency in AUCs of radiomics signature in the training, internal validation, and external validation cohorts were 0.907 (95% CI: 0.857-0.957), 0.771 (95% CI: 0.682-0.845) and 0.813 (95% CI: 0.730-0.878), respectively (Figure S2).

**DLS Construction and Analysis**
Initially, 3,337 DL features displayed significant differences between the MIA and IAC groups. After using the standard of minimum redundancy maximum relevance, 501 DL features were retained. 18 DL features with non-zero coefficients were finally selected to be included in the DLS score calculation formula via LASSO analysis (Figure 2). Details of the selected DL features are shown in Supplementary Table S1. The DLS scores between the MIA and IAC groups were statistically different in the training, internal validation, and external validation cohorts (all $P<0.001$; Table 1). Details of calculation information for DLS score are shown in Supplementary A5.

Convolutional filters from the DLS were pictured to facilitate the comprehending of the DL features [22]. Specifically, convolution filters from shallow layers may provide information on low-level uncomplicated features, such as shape-dependent features (first column of Figure 3A). More complex features of tumor shape, such as circular structure or an arch shape could be learned from a deeper convolutional layer (second column of Figure 3A). As the layer deepened, more abstract features were found to accurately differentiate SSPNs (third column of Figure 3A). These DL features were not explicable by visually, but tended to be related to the invasiveness degree of SSPNs.

To further investigate the connection between the DL features of the SSPNs and their final diagnoses, a negative filter and positive filter (Figure 3B) were extracted from the 18 features screened by LASSO, which could capture different texture patterns. The negative filter reacted intensely when MIA lesions were entered into the DLS, whereas the positive filter reacted strongly when IAC lesions were included. Clearly diverse clusters of response could consequently be detected in lesion images of MIA and IAC.
**DLN Construction and Validation**

Two independent predictors were identified in the multivariable logistic regression analysis, which were the DLS score (OR: 4.287, 95% CI: 2.433-7.554; \( P<0.001 \)) and the lobulated sharp (OR: 2.623, 95% CI: 1.005-6.849; \( P=0.049 \)). We constructed the DLN model by incorporating these two factors (Table 2; Figure 4).

The diagnostic performance of the subjective model, DLS, and DLN is displayed in Table 3 and is detailed in Figure 5. The diagnostic accuracy of DLN (AUC=0.889, 95% CI: 0.824-0.936) for SSNPs was higher than that of the DLS (AUC=0.881, 95% CI: 0.815-0.930) and the subjective model (AUC=0.819, 95% CI: 0.744-0.880) in the training cohort. The results were confirmed by an internal validation cohort (AUC=0.915, 95% CI: 0.846-0.959 for the DLN; AUC=0.882, 95% CI: 0.796-0.927 for the DLS; AUC=0.811, 95% CI: 0.726-0.879 for the subjective model) and an external validation cohort (AUC=0.914, 95% CI: 0.848-0.958 for the DLN; AUC=0.907, 95% CI: 0.840-0.953 for the DLS; AUC=0.782, 95% CI: 0.697-0.853 for the subjective model).

The results of Delong test demonstrated significant differences of AUCs between the subjective model and the DLN in the training, internal validation, and external validation cohorts (\( P=0.009 \), \( P=0.011 \), and \( P=0.002 \), respectively). No significant differences of AUCs between the DLS and the DLN in the training, internal validation, and external validation cohorts were observed (\( P=0.441 \), \( P=0.083 \), and \( P=0.611 \), respectively).

NRI analysis showed that the DLN had significantly better diagnostic performance than the subjective model and the DLS in the training (NRI=0.498, \( P=0.003 \); NRI=1.039, \( P<0.001 \)), internal validation (NRI=0.258, \( P=0.004 \);
NRI=1.007, $P<0.001$), and external validation cohorts (NRI=0.618, $P<0.001$; NRI=1.183, $P<0.001$), respectively.

After stratified analysis, we found that the performance of the DLN was not altered by patient gender, age, CT scan equipment, slice thickness and nodule size (all $P>0.05$; Supplementary A6 and Figure S3). DCA demonstrated a higher net advantage for the DLN in differentiating the MIA from IAC groups than the DLS and the subjective model. The threshold probability ranged from 0.03 to 0.95 (Figure 6). These results indicate that the DLN displayed better diagnostic efficiency for differentiating MIA from IAC compared with the subjective model and the DLS.

Discussion

This study assessed the value of the DL method in the preoperative diagnosis of SSPNs. Our results demonstrated that the proposed DLN model had better diagnostic performance than the DLS and the subjective models in the discrimination of MIA from IAC.

Subjective CT manifestations have been studied to identify the invasiveness degree of SSPNs in previous literature, which noted that SSPNs with solid components are more likely to be invasive lesions [6, 7]. However, more than one-third of SSPNs are pathologically diagnosed as pre-invasive or MIA lesions after surgery [23]. Previous studies have shown that nodule diameter of more than 10 mm is an optimal critical value for distinguishing invasive lesions from non-invasive lesions [24]. Likewise, the cut-off value of 10 mm performed well for SSPNs in our study. Research has also confirmed the morphologic CT feature of lobulated sharp as a useful predictor for distinguishing IAC from SSPNs [7]. Although the radiological characteristics of SSPNs can be easily defined, researchers have noted that inevitable
inter- and intra-observer variability are prevalent among radiologist interpretations [9]. Thus, a more objective pragmatic definition is warranted to accurately differentiate MIA from IAC.

Radiomic signatures reflect intratumoral heterogeneity and are associated with gene-expression profiles [12, 25]. Recent studies have reported that radiomics-based nomogram can be applied as a non-invasive biomarker for evaluating the invasion degree of SSPNs [26, 27]. However, clinical application of radiomics is still limited due to several barriers, such as lack of easily manipulated tools and variability of radiomic features [28]. Compared with the conventional radiomics methods, DL is an end-to-end data-driven method, which automatically extracts task related features from raw data and generate suitable models for tasks. Meanwhile, the DL features can detail various levels of tumor information from low-level characteristics by visual to high-level abstract features, such as edge related feature, circular structure, arch shape and more abstract features with the invasive status of SSPNs (Figure 3).

Although there has been great improvement in explaining the machine learning/DL system, a comprehensive understanding of the internal mechanism in the deep neural network model is still a major task. Most clinical studies on the DL model have been conducted in a single institution or with a limited subject number [17, 18]. Yanagawa et al compared the predictive ability of pathological invasiveness in lung adenocarcinoma between radiologists and a DL system, which showed that the diagnostic performance of DL (AUC: 0.712) was similar to the experienced radiologist (AUC: 0.714). However, the sample size in the study (n=90) was relatively small [18]. Zhao et al observed 560 patients with 651 nodules to evaluate the added value of DL features to routine visual assessment. They found the DL system achieved better classification performance than the radiologists in automatic
invasiveness prediction (3D Dense-Sharp Model, AUC: 0.880; 3D Dense-Net Model, AUC: 0.874). Whereas, both subsolid and solid nodules were enrolled in this study, which could influence the pathological diagnosis in clinical practice by interfering nodule subtypes [17]. For the first time, we studied the DLN on the combination of DLS and subjective CT findings. Because deep learning features can decode the phenotypes of SSPNs, the characteristics of lesions were comprehensively and quantitatively learned by the fusion DLN model. Moreover, our study differed from the previous studies in terms of generalizability. The diagnostic ability of the DLN model was superior to the subjective model, and significant differences were found between the two models in the training, internal validation, and external validation cohorts (DeLong test: $P=0.009$, $P=0.011$, $P=0.002$, respectively).

This study had several limitations. First, as a retrospective study, potential selection bias may exist; thus, prospective studies with larger cohorts are needed to verify the reproducibility and generalizability of the DLN model. Second, chest CT images were conducted using non-uniform scanning equipment. This may have resulted in the divergence of CT attenuation values with resultant bias in DL features. However, no significant differences were detected in stratified analysis in regard to CT scan equipment, slice thickness and nodule size. Third, only patients who have received surgery and were pathologically confirmed to be lung adenocarcinomas were included, these SSPNs have a higher probability to be malignant on chest CT scan. Patients with pre-invasive lesions such as AAH (atypical adenomatous hyperplasia) and AIS (adenocarcinoma in situ) were excluded, as they do not need immediate therapeutic intervention in clinical practice. Finally, since the evaluation of
intraoperative frozen section (FS) can provide additional diagnostic information, incorporation of this index in our DLN model may further improve the predictive value in judging the invasion status of SSPNs. In the future, we will collect more patients with FS pathological data to explore this issue.

Conclusions

In this study, we combined the DLS and subjective CT parameters to construct the DLN, which was a non-invasive, quantitative, and reproducible model that could be used to differentiate MIA from IAC sensitively and specifically. Our DLN model may improve the diagnostic accuracy of SSPNs and facilitate the planning of the most appropriate treatment strategy.

References


[8] S. Lee, C. Park, J. Goo, H. Lee, J. Wi, C. Kang, Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules:


nODULES, LUNG CANCER 108 (2017) 192-197,

https://doi.org/10.1016/j.lungcan.2017.03.011.


https://doi.org/10.2214/AJR.18.20623.


https://doi.org/10.1148/radiol.2018172361.


https://doi.org/10.1038/nature14539.


[19] M. Fischer Andreas, V. Akos, A. van, G. Parkwood, S. Pooyan, I. Sperl Jonathan, W. Nance John, S. Joseph, Comparison of artificial intelligence-based fully automatic chest CT emphysema quantification to pulmonary function testing,
American journal of roentgenology 214 (2020) 1065-1071,
https://doi.org/10.2214/AJR.19.21572.


**Figure Legends.**

**Figure 1.**
Deep learning nomogram workflow. (A) CT images process of SSPNs; (B) Deep learning features extraction; (C) DLS construction; (D) DLN establishment; (E) Diagnostic models evaluation and comparison. SSPNs: subsolid pulmonary nodules; CT: computed tomography; MIA: minimally invasive adenocarcinoma; IAC: invasive adenocarcinoma; ROI: region of interest; DLS: deep learning signature; DLN: deep learning nomogram.

**Figure 2.**
Deep learning features extraction and selection process. (A) The tuning parameter ($\lambda$) was screened using 10-fold cross-validation based upon minimum criteria. $\log(\lambda)$ was plotted on the X-axis, and binomial deviance was plotted on the Y-axis. The dotted vertical lines demonstrate optimal values determined by the minimum criterion and one standard error of the minimum criterion (1-SE). Optimal $\lambda=0.02086$; $\log(\lambda)=-3.8699$. (B) LASSO coefficient profiled of whole features. Coefficient profiles were plotted against $\log(\lambda)$. The vertical line was drawn where the 18 optimal radiomics features with non-zero coefficients were indicated in the plot according to 10-fold cross-validation.

**Figure 3.**
Deep learning analysis. (A) Convolutional filters from the deep learning model. Each convolutional layer included hundreds of convolutional filters, and three convolution filters were chosen to be illustrated in each layer. (B) Response with the positive and negative filters in IAC and MIA. The positive and negative filters had strong responses to IAC and MIA nodules, respectively. MIA: minimally invasive adenocarcinoma; IAC: invasive adenocarcinoma.
Figure 4.

Construction of deep learning nomogram. Calibration curves of the deep learning radiomics nomogram in the training (A), internal validation (B), and external validation cohorts (C), respectively. DLS: deep learning signature.

Figure 5.

ROC curves of the subjective model (black), DLS (red), and DLN (green) in the training (A), internal validation (B), and external validation (C) cohorts. DLS: deep learning signature; DLN: deep learning nomogram; ROC: receiving operating characteristic; AUC: area under curve; CI: confidence interval.

Figure 6.

Decision-curve analysis. The net advantage of each model was plotted on the Y-axis. The X-axis demonstrated the threshold probability, which was calculated as when the expected advantage of receiving treatment was the same to that of avoiding treatment. The subjective model, DLS, and DLN are illustrated by the black, red, and green dotted lines, respectively. The gray and black solid lines indicate the assumptions that all or no patients had IAC nodules, respectively. The DLN provided the highest net advantage (threshold probability: 0.03-0.95). DLS: deep learning signature; DLN: deep learning nomogram; IAC: invasive adenocarcinoma.
### TABLE 1: Clinicopathological Characteristics and CT Imaging Assessment of MIA and IAC in Patients with SPSNs

<table>
<thead>
<tr>
<th></th>
<th>Training Cohort (n=136)</th>
<th></th>
<th></th>
<th></th>
<th>Internal Validation Cohort (n=111)</th>
<th></th>
<th></th>
<th></th>
<th>External Validation Cohort (n=118)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MIA (n=58)</td>
<td>IAC (n=78)</td>
<td>P</td>
<td>MIA (n=44)</td>
<td>IAC (n=67)</td>
<td>P</td>
<td>MIA (n=48)</td>
<td>IAC (n=70)</td>
<td>P</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>15</td>
<td>30</td>
<td>0.122</td>
<td>18</td>
<td>30</td>
<td>0.687</td>
<td>9</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>43</td>
<td>48</td>
<td>&lt;0.001*</td>
<td>26</td>
<td>37</td>
<td>0.009*</td>
<td>39</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age (median± SD; years)</strong></td>
<td>51.19±12.13</td>
<td>57.99±8.89</td>
<td></td>
<td>51.91±12.26</td>
<td>57.69±11.42</td>
<td></td>
<td>52.98±10.09</td>
<td>59.24±9.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUL</td>
<td>23</td>
<td>28</td>
<td></td>
<td>28</td>
<td>9</td>
<td></td>
<td>14</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LLL</td>
<td>5</td>
<td>9</td>
<td></td>
<td>6</td>
<td>6</td>
<td></td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RUL</td>
<td>18</td>
<td>27</td>
<td>0.954</td>
<td>19</td>
<td>26</td>
<td>0.399</td>
<td>13</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RML</td>
<td>4</td>
<td>5</td>
<td></td>
<td>2</td>
<td>4</td>
<td></td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RLL</td>
<td>8</td>
<td>9</td>
<td></td>
<td>8</td>
<td>7</td>
<td></td>
<td>7</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-solid</td>
<td>51</td>
<td>39</td>
<td>&lt;0.001*</td>
<td>38</td>
<td>29</td>
<td>&lt;0.001*</td>
<td>31</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-solid</td>
<td>7</td>
<td>39</td>
<td></td>
<td>6</td>
<td>38</td>
<td></td>
<td>17</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Size (mm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 10</td>
<td>40</td>
<td>15</td>
<td>&lt;0.001*</td>
<td>33</td>
<td>16</td>
<td>&lt;0.001*</td>
<td>38</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥10</td>
<td>18</td>
<td>63</td>
<td></td>
<td>11</td>
<td>51</td>
<td></td>
<td>10</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Margin</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular</td>
<td>36</td>
<td>19</td>
<td>&lt;0.001*</td>
<td>27</td>
<td>9</td>
<td>&lt;0.001*</td>
<td>33</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irregular</td>
<td>22</td>
<td>59</td>
<td></td>
<td>17</td>
<td>58</td>
<td></td>
<td>15</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lobulated Sharp</strong></td>
<td>12</td>
<td>51</td>
<td>&lt;0.001*</td>
<td>13</td>
<td>49</td>
<td>&lt;0.001*</td>
<td>7</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td>46</td>
<td>27</td>
<td></td>
<td>31</td>
<td>18</td>
<td></td>
<td>41</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Air-bronchogram Sign</strong></td>
<td>1</td>
<td>18</td>
<td>&lt;0.001*</td>
<td>2</td>
<td>17</td>
<td>0.004*</td>
<td>4</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence</td>
<td>57</td>
<td>60</td>
<td></td>
<td>42</td>
<td>50</td>
<td></td>
<td>44</td>
<td>67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deep Learning Signature Score</strong> (median± SD)</td>
<td>0.35±0.22</td>
<td>0.74±0.18</td>
<td>&lt;0.001</td>
<td>0.40±0.26</td>
<td>0.66±0.24</td>
<td>&lt;0.001</td>
<td>0.44±0.23</td>
<td>0.71±0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Radiomics Signature Score</strong> (median± SD)</td>
<td>0.37±0.23</td>
<td>0.72±0.17</td>
<td>&lt;0.001</td>
<td>0.39±0.18</td>
<td>0.65±0.14</td>
<td>&lt;0.001</td>
<td>0.52±0.20</td>
<td>0.81±0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The differences between MIA and IAC were analyzed by Wilcoxon rank-sum test or Pearson’s chi-squared test. MIA = minimally invasive adenocarcinoma, IAC = invasive adenocarcinoma, SPSN = subsolid pulmonary nodule, LUL = left upper lobe, LLL = left lower lobe, RUL = right upper lobe, RML = right middle lobe, RLL = right lower lobe. *P-value < 0.05.

### TABLE 2: Multivariate Logistic Regression Analysis of Parameters for Distinguishing between MIA and IAC

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>Odds Ratio (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.537</td>
<td>0.585</td>
<td>0.085</td>
</tr>
<tr>
<td>Lobulated Sharp</td>
<td>0.964</td>
<td>2.623 (1.005-6.849)</td>
<td>0.049*</td>
</tr>
<tr>
<td>Deep Learning Signature</td>
<td>1.456</td>
<td>4.287 (2.433-7.554)</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

MIA: minimally invasive adenocarcinoma; IAC: invasive adenocarcinoma; β: regression coefficient; CI: confidence interval. *P-value < 0.05.
TABLE 3: Diagnostic Performance of the Models in the Training, Internal Validation, and External Validation Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Training Cohort (n=136)</th>
<th></th>
<th>Internal Validation Cohort (n=111)</th>
<th></th>
<th>External Validation Cohort</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subjective Model</td>
<td>Deep Learning Signature</td>
<td>Deep Learning Nomogram</td>
<td>Subjective Model</td>
<td>Deep Learning Signature</td>
<td>Deep Learning Nomogram</td>
</tr>
<tr>
<td></td>
<td>(90x742)</td>
<td>(91x673)</td>
<td>(91x625)</td>
<td>(90x673)</td>
<td>(91x625)</td>
<td>(90x673)</td>
</tr>
<tr>
<td>AUC</td>
<td>0.819</td>
<td>0.881</td>
<td>0.889</td>
<td>0.811</td>
<td>0.882</td>
<td>0.915</td>
</tr>
<tr>
<td>(9x730)</td>
<td>(0.744-0.880)</td>
<td>(0.815-0.930)</td>
<td>(0.824-0.936)</td>
<td>(0.726-0.879)</td>
<td>(0.796-0.927)</td>
<td>(0.846-0.959)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.859</td>
<td>0.846</td>
<td>0.897</td>
<td>0.731</td>
<td>0.851</td>
<td>0.970</td>
</tr>
<tr>
<td>(9x649)</td>
<td>(67/78)</td>
<td>(66/78)</td>
<td>(70/78)</td>
<td>(49/67)</td>
<td>(57/67)</td>
<td>(65/67)</td>
</tr>
<tr>
<td></td>
<td>0.672</td>
<td>0.759</td>
<td>0.707</td>
<td>0.796</td>
<td>0.750</td>
<td>0.750</td>
</tr>
<tr>
<td>(9x619)</td>
<td>(39/58)</td>
<td>(44/58)</td>
<td>(41/58)</td>
<td>(35/44)</td>
<td>(33/44)</td>
<td>(33/44)</td>
</tr>
<tr>
<td></td>
<td>0.779</td>
<td>0.809</td>
<td>0.816</td>
<td>0.757</td>
<td>0.811</td>
<td>0.883</td>
</tr>
<tr>
<td>(9x673)</td>
<td>(106/136)</td>
<td>(110/136)</td>
<td>(111/136)</td>
<td>(84/111)</td>
<td>(90/111)</td>
<td>(98/111)</td>
</tr>
<tr>
<td></td>
<td>0.779</td>
<td>0.825</td>
<td>0.805</td>
<td>0.845</td>
<td>0.838</td>
<td>0.855</td>
</tr>
<tr>
<td>(9x625)</td>
<td>(67/86)</td>
<td>(66/80)</td>
<td>(70/87)</td>
<td>(49/58)</td>
<td>(57/68)</td>
<td>(65/76)</td>
</tr>
<tr>
<td></td>
<td>0.780</td>
<td>0.786</td>
<td>0.837</td>
<td>0.660</td>
<td>0.767</td>
<td>0.943</td>
</tr>
<tr>
<td>(9x625)</td>
<td>(39/50)</td>
<td>(44/56)</td>
<td>(41/49)</td>
<td>(35/53)</td>
<td>(33/43)</td>
<td>(33/35)</td>
</tr>
</tbody>
</table>

Note-AUC= area under the curve, CI= confidence interval, PPV= positive predictive value, NPV= negative predictive value.
Highlights

a. The deep learning nomogram was developed to preoperatively differentiate MIA from IAC in patients with SSPNs.

b. The deep learning nomogram achieved superior performance compared to the deep learning signature, or the clinical model alone.

c. The deep learning nomogram is an end-to-end data-driven method and directly predicts the status of SSPNs without further human assistance.
Author Statement:

**Xiangmeng Chen:** Conceptualization, Methodology, Investigation, Data curation, Roles/Writing – original draft, Writing – review & editing

**Bao Feng:** Conceptualization, Methodology, Visualization, Roles/Writing – original draft, Writing – review & editing

**Yehang Chen:** Software, Visualization, Roles/Writing – original draft, Writing – review & editing

**Xiaobei Duan:** Formal analysis, Validation, Roles/Writing – original draft, Writing – review & editing

**Kunfeng Liu:** Investigation, Roles/Writing – original draft, Writing – review & editing

**Kunwei Li:** Investigation, Roles/Writing – original draft, Writing – review & editing

**Chaotong Zhang:** Project administration, Roles/Writing – original draft, Writing – review & editing

**Xueguo Liu:** Project administration, Supervision, Data curation, Roles/Writing – original draft, Writing – review & editing

**Wansheng Long:** Project administration, Supervision, Data curation, Resources, Roles/Writing – original draft, Writing – review & editing