3D Deep Learning Versus the Current Methods for Predicting Tumor Invasiveness of Lung Adenocarcinoma

Yilv Lv  
Shanghai Chest Hospital: Shanghai Jiao Tong University Affiliated Chest Hospital

Ying Wei  
Shanghai United Imaging Intelligence Co., Ltd.

Kuan Xu  
Shanghai Chest Hospital: Shanghai Jiao Tong University Affiliated Chest Hospital

Xiaobin Zhang  
Shanghai Chest Hospital: Shanghai Jiao Tong University Affiliated Chest Hospital

Rong Hua  
Shanghai Chest Hospital: Shanghai Jiao Tong University Affiliated Chest Hospital

Jia Huang  
Shanghai Chest Hospital: Shanghai Jiao Tong University Affiliated Chest Hospital

Min Li  
Shanghai Municipal Hospital of Traditional Chinese Medicine

Cui Tang  
Tongji University Affiliated Yangpu Hospital: Shanghai Yangpu District Central Hospital

Long Yang  
Gansu Medical College

Bingchun Liu  
Weifang People's Hospital

Yonggang Yuan  
Qilu Hospital of Shandong University

Siwen Li  
Qingyuan People's Hospital

Yaozong Gao  
Shanghai United Imaging Intelligence Co., Ltd.

Xianjie Zhang  
Shanghai United Imaging Intelligence Co., Ltd.

Yifan Wu  
Shanghai United Imaging Intelligence Co., Ltd.

Yuchen Han
Research Article

**Keywords:** computer-aided diagnosis, lung adenocarcinoma, intraoperative frozen section, tumor invasiveness

**Posted Date:** March 2nd, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1377789/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Different pathological subtypes of lung adenocarcinoma lead to different treatment decisions and prognoses, and it is clinically important to distinguish invasive lung adenocarcinoma from preinvasive adenocarcinoma (adenocarcinoma in situ and minimally invasive adenocarcinoma). This study aims to investigate the performance of the deep learning approach on the classification of tumor invasiveness and compare it with the performances of currently available approaches.

**Methods:** In this study, we propose a deep learning approach based on 3D conventional networks to automatically predict the invasiveness of pulmonary nodules. A total of 901 early-stage non-small cell lung cancer patients who underwent surgical treatment at Shanghai Chest Hospital between November 2015 and March 2017 were retrospectively included and randomly assigned to a training set (n=814) or testing set 1 (n=87). We subsequently included 116 patients who underwent surgical treatment and intraoperative frozen section between April 2019 and January 2020 to form testing set 2. We compared the performance of our deep learning approach in predicting tumor invasiveness with that of intraoperative frozen section analysis and human experts (radiologists and surgeons).

**Results:** The deep learning approach yielded an area under the receiver operating characteristic curve (AUC) of 0.946 for distinguishing preinvasive adenocarcinoma from invasive lung adenocarcinoma in testing set 1, which is significantly higher than the AUCs of human experts (P<0.05). In testing set 2, the deep learning approach distinguished invasive adenocarcinoma from preinvasive adenocarcinoma with an AUC of 0.862, which is higher than that of frozen section analysis (0.755, P=0.043), senior thoracic surgeons (0.720, P=0.006), radiologists (0.766, P>0.05) and junior thoracic surgeons (0.768, P>0.05).

**Conclusions:** We developed a deep learning approach that achieved comparable performance to intraoperative frozen section analysis in determining tumor invasiveness. The proposed method may contribute to clinical decisions related to the extent of surgical resection.

1. Introduction

Lung cancer ranks second in the most commonly diagnosed cancer and remains the leading cause of cancer death worldwide (1, 2). With the widespread implementation of low-dose computed tomography (CT) screening and regular physical examinations, a substantial number of early-stage lung cancers have been detected (3). Surgical resection remains the gold standard for early-stage lung cancer treatment, and the mode of surgery is lobectomy (4). However, an increasing number of studies and single-institution trials have demonstrated that sublobar resection can achieve comparable outcomes in well-selected patients with early-stage non-small cell lung cancer (NSCLC) (5, 6). Sublobar resection can preserve the lung parenchyma, which is particularly valuable for patients with poor pulmonary reserve or those who are likely to require subsequent additional resection (5). Therefore, sublobar resection is vital in the treatment of patients with early-stage NSCLC.
A consistent approach to identify optimal candidates with a low likelihood of recurrence for sublobar resection of NSCLC has not been established. Patients with ground-glass opacity-dominant clinical stage IA adenocarcinomas are suitable for sublobar resection, as confirmed by the latest clinical trial (7). In the new multidisciplinary classification of pulmonary adenocarcinoma by the International Association for the Study of Lung Cancer (IASLC)/American Thoracic Society (ATS)/European Respiratory Society (ERS), the presence of adenocarcinoma in situ (AIS) and minimally invasive adenocarcinoma (MIA) are associated with 100% or near-100% disease-specific survival, respectively, after complete resection. Invasive lung adenocarcinoma (IAC) is more aggressive and has a poorer prognosis than AIS and MIA, suggesting that sublobar resection is only appropriate for patients with MIA or AIS (8, 9).

Currently, there are three modalities to evaluate pathological invasiveness and the suitability of sublobar resection in patients with early lung adenocarcinoma: preoperative biopsy, CT imaging, and intraoperative frozen section analysis. Small lesions are challenging to locate, while biopsy samples may not be representative (10, 11). Additionally, it remains controversial whether preoperative biopsy enhances the possibility of early lung cancer recurrence (12, 13). Intraoperative frozen section analysis has traditionally been used to assess tumor invasiveness and guide surgical management. However, the technique does have certain limitations: Several studies have shown that the accuracy and sensitivity of intraoperative frozen sections are relatively low for subcentimeter pulmonary nodules (14, 15). There has been a strong focus on identifying pathological invasiveness according to imaging findings. CT imaging can reportedly distinguish preinvasive lung adenocarcinoma (pre-IAC; AIS and MIA) from IAC, although the small sample sizes and ambiguous appearances of these findings prevent its routine adoption in clinical practice (16–20). It is therefore a great challenge for radiologists or experts to diagnose a large number of detected pulmonary nodules, as these methods are time consuming and error-prone when interpreting nodules. Therefore, we need a more straightforward and precise method to determine the pathological aggressiveness of all types of nodules based on CT imaging, not just ground-glass nodules.

In recent years, artificial intelligence (AI) techniques coupled with radiological imaging have played an essential role in automatically predicting the tumor invasiveness of pulmonary adenocarcinomas from CT scans (21–25). Deep learning, a popular research area of AI, enables end-to-end models to obtain self-learned features and achieves promising results using input data without the need for manual feature extraction (26). Deep learning algorithms have been widely applied to many problems, such as lung nodule detection, segmentation, and classification (27, 28).

The purpose of this study was to develop a computer-aided approach to accurately and automatically discriminate the invasiveness of lung adenocarcinomas in routine chest CT images. We built a deep learning approach and investigated the utility of the model in predicting pathological invasiveness among patients with early-stage lung adenocarcinoma. In addition, we compared the performance of the deep learning model with that of observers and intraoperative frozen section diagnoses to determine the best method of distinguishing pre-IAC from IAC in clinical practice.

2. Methods
2.1. Ethical considerations

This retrospective study adhered to the Declaration of Helsinki and relevant ethical policies in China. The study protocol was approved by the Institutional Review Board and Ethics Committee of Shanghai Chest Hospital (No. IS2180). The requirement for patient consent was waived because of the retrospective study design.

2.2. Data collection

This study retrospectively reviewed the medical records of 2671 consecutive patients with NSCLC who underwent surgical resection in Shanghai Chest Hospital between November 2015 and March 2017 to develop the training set and testing set 1. An additional dataset of 273 patients who underwent surgery between April 2019 and January 2020 was separately identified and formed an additional testing set (i.e., testing set 2). The inclusion criteria were as follows: (1) stage 0 or IA lung adenocarcinoma confirmed by final pathology according to the 8th Edition of the TMN Classification (29); (2) availability of preoperative thin-section CT (0.625 mm–1.25 mm) images; and (3) resected nodules were sent for paraffin sectioning, and the final pathological results were available. The exclusion criteria were as follows: (1) previous history of malignant tumor, radiation therapy, chemotherapy, or biopsy; (2) other malignant diseases; (3) insufficient CT or pathology quality; and (4) multiple pulmonary nodules. Finally, 901 patients with early-stage lung adenocarcinoma were enrolled and randomly divided into the training and validation sets at a ratio of 8:1, and 116 patients were enrolled in the independent testing set. To compare the accuracy of intraoperative frozen section analysis with that of artificial intelligence-based CT image analysis, frozen section diagnoses of the independent testing set 2 were collected (Fig. 1).

2.3. CT image acquisition, classification, and pathological evaluation

All preoperative CT scans were obtained at full inspiration to avoid respiratory motion artifacts. Brilliance iCT and Ingenuity (Philips Medical Systems, Netherlands) scanners were used to scan CT images at an efficient dose of 120 kV tube energy and 200 mA. All CT data were acquired in the supine position at full inspiration. High-resolution images were acquired with a reconstruction slice thickness of 1 mm and no overlap, and a lung window (window width: 1500, window level: -500) was used for film reading.

For frozen section diagnosis, resected tumor tissues were preserved in a sterile, sealed plastic bag; they were sent to the pathology department within 5 min after resection. Essential tumor information was recorded; one block of the largest tumor tissue was separated from the sample and sectioned using a CM-3050s freezing microtome (Leica, Nussloch, Germany). Before sectioning, the tissue block was frozen at -24°C for 5 min in OCT compound (SAKURA Tissue-Tek; Torrance, CA, USA). One or two slices (5 µm each) were collected and placed on glass slides. The slides were fixed in methanol/glacial acetic acid for 10–20 s and then subjected to routine hematoxylin and eosin staining (Fig. 2). The predominant pattern was defined according to the histologic component with the greatest percentage.
For paraffin-embedded sections, any remaining tissues that had been collected during surgery were fixed in 10% formaldehyde, embedded in paraffin, continuously sectioned at 5 µm, and subjected to hematoxylin and eosin staining for postoperative pathological analysis. Final pathology was also established via elastic fiber staining and immunohistochemical assessment of cytokeratin 7, thyroid transcription factor-1, and napsin A (all antibodies from Cell Signaling Technology; Danvers, MA, USA) in paraffin-embedded sections.

Frozen section and final pathology diagnoses came from blind assessments by two pathologists (Y.H. and Z.S., chest pathologists with more than 20 years of experience in pathological diagnosis) according to the IASLC/ATS/ERS classification (8). Two pathologists reevaluated the diagnoses to reach a consensus if discrepancy presented. AIS and MIA were combined to form a low-risk group that was called pre-IAC.

2.4. Nodule labeling and segmentation

All lung nodules with nodule diameters greater than 3 mm on each CT scan were automatically localized with 3D bounding boxes and automatically segmented using a research tool (30) developed by Shanghai United Imaging Intelligence Co., Ltd. A total of 1017 nodules were ultimately included as regions of interest (ROIs), and each of them was reviewed and confirmed by at least two senior radiologists. Supplementary Fig. 1 illustrates the size distribution of pre-IAC and IAC nodules on diameter.

2.5. Deep learning model construction

In the data preprocessing step, each CT image was clipped to the range of -400–1500 HU to remove most of the variations induced by bone structures. We then normalized the intensity of images into the range of 0 to 1 before feeding them into the deep learning network. To avoid overfitting and increase the robustness of the deep learning network, image augmentation, including rotation, scaling, and flipping, was performed on each image with a probability of 0.5. Rotation was randomly performed with an angle along an axis in a range of −5° to 5°. The scaling factor was randomly sampled in a range from 0.75 to 1.25. Flipping was adopted randomly along each axis.

The deep learning model was built by using a convolutional neural network (CNN), which consists of one input block, four downsample blocks, and one output block (Fig. 1). The input block includes a 3D convolution layer with a kernel size of 3, and the downsample block includes a 3D convolution layer with a kernel size of 3 and stride of 2. Each convolution is followed by a batch normalization layer and ReLU layer. In the output block, two fully connected layers are used to represent the final output features. These two fully connected layers are followed by a ReLU layer and a softmax layer. Finally, the output layer includes two nodes as the predicted probabilities for pre-IAC and IAC. Note that the deep learning model used only the image information where clinical features were not included.

2.6. Subcentimeter nodule classification model construction
Considering that small nodules are more difficult to discriminate than nodules with larger sizes, we collected subcentimeter nodules with sizes no greater than 10 mm from the training set, testing set 1, and testing set 2. We then trained a specific model on the subcentimeter nodules of the training set, with the same training strategies used for deep learning model construction. The performance was evaluated on testing set 1 and testing set 2 (Fig. 1).

2.7. Observer study

For human performance comparisons, two radiologists, two junior surgeons, and two senior surgeons were recruited. They were blinded to the clinical records and pathological results and diagnosed all the nodules with only CT images. Each reader read the CT images independently and classified the nodules into pre-IAC or IAC, as with the deep learning model.

2.8. Statistical analysis

Age, sex, smoking history, surgical procedure, tumor size, and location for each patient were analyzed. Pearson’s $\chi^2$ test or Fisher’s exact test was used to compare frequencies of categorical variables (all continuous variables were converted to categorical variables except for age, as shown in Table 1). The Mann-Whitney U test was used to analyze the age between two groups. The diagnostic performance of artificial intelligence models, observers, and frozen section diagnoses was evaluated by the area under the receiver operating characteristic (ROC) curve (AUC) and other evaluation metrics, such as accuracy, sensitivity, and specificity. The DeLong test was performed to compare the AUC curves of the deep learning models and observer studies and intraoperative frozen section, and the 95% confidence interval (95% CI) of the AUC was also assessed. In addition, the statistical significance of the difference in accuracy between deep learning models, observers, and frozen section diagnoses was evaluated using Pearson’s $\chi^2$. All statistical analyses reported in this study were performed with Python (Version 3.7.0) and R (Version 4.0.2), and a p-value less than 0.05 was considered statistically significant.

| Final Pathology | Overall nodules | Subcentimeter nodules (< 10 mm) |
|-----------------|-----------------|---------------------------------|
|                 | Training set    | Testing set 1 | Testing set 2 | Total | Training set 1 | Testing set 1 | Testing set 2 | Total |
| AIS/MIA         | 339             | 35            | 48            | 422   | 272             | 30            | 33            | 335   |
| IAC             | 475             | 52            | 68            | 595   | 45              | 5             | 15            | 65    |
| Total           | 814             | 87            | 116           | 1017  | 317             | 35            | 48            | 400   |

Abbreviations: AIS = adenocarcinoma in situ, MIA = minimally invasive adenocarcinoma, IAC = invasive lung adenocarcinoma.

3. Results
A total of 1017 nodules (pre-IAC/IAC: 422/595) were included. The number of nodules for the overall and subcentimeter nodule classification is shown in Supplementary Table 1, and the specific clinicopathological characteristics are summarized in Table 1. Significant differences were found in terms of age, sex, smoking history, nodule diameter, and surgical type in the primary dataset (P < 0.05). There were also significant differences between AIS/MIA and IAC in terms of age, nodule diameter, and surgical type in the independent testing set (P < 0.05).

3.1. Classification performance on overall nodule size

The deep learning model was trained on over 500 epochs, and after convergence, the weights were used for testing. The results show that the deep learning model achieved an AUC of 0.946 (95% CI: 89.9–99.4) with a sensitivity of 86.5%, specificity of 91.4%, and accuracy of 88.5% on testing set 1 and an AUC of 0.862 (95% CI: 79.4–93.0) with a sensitivity of 73.5%, specificity of 91.7%, and accuracy of 81.0% on testing set 2 (Table 2).
Table 2
The performance of overall nodules with various methods for predicting pathological invasiveness

|                  | F1-score | Sensitivity | Specificity | MCC  | ACC  | AUC                  |
|------------------|----------|-------------|-------------|------|------|----------------------|
| **Testing set 1**|          |             |             |      |      |                      |
| Deep Learning Model | 0.900    | 0.865       | 0.914       | 0.769 | 0.885 | 0.946 [0.899, 0.994] |
| Radiologists*     | 0.869    | 0.913       | 0.714       | 0.666 | 0.833 | 0.809 [0.711, 0.907] |
| Thoracic Surgeon (Junior)* | 0.813    | 0.724       | 0.913       | 0.624 | 0.799 | 0.823 [0.728, 0.918] |
| Thoracic Surgeon (Senior)** | 0.740    | 0.615       | 0.929       | 0.546 | 0.741 | 0.779 [0.675, 0.883] |
| **Testing set 2** |          |             |             |      |      |                      |
| Frozen Section*   | 0.820    | 0.676       | 0.833       | 0.624 | 0.741 | 0.755 [0.663, 0.847] |
| Deep Learning Model | 0.820    | 0.735       | 0.917       | 0.644 | 0.810 | 0.862 [0.794, 0.930] |
| Radiologists      | 0.818    | 0.829       | 0.720       | 0.558 | 0.784 | 0.776 [0.687, 0.865] |
| Thoracic Surgeon (Junior) | 0.759    | 0.662       | 0.885       | 0.544 | 0.754 | 0.768 [0.678, 0.858] |
| Thoracic Surgeon (Senior)* | 0.675    | 0.551       | 0.896       | 0.463 | 0.694 | 0.720 [0.623, 0.817] |

* Significant difference found between this diagnostic method of AUC and deep learning model by the DeLong test (P < 0.05).

** Significant difference found between this diagnostic method of accuracy and the deep learning model by Pearson's χ² test (P < 0.05).

### 3.2. Classification performance on subcentimeter nodule size

The results of nodules with sizes no greater than 10 mm are shown in Table 3. The results show that the deep learning model achieved an AUC of 0.893 (95% CI: 77.2–100.0) on testing set 1. On testing set 2, the deep learning model achieved a sensitivity of 40.0%, specificity of 97.0%, accuracy of 85.7%, and AUC of 0.646 (95% CI: 42.9–86.4).
Table 3
The performance of subcentimeter nodules (< 10 mm) with various methods for predicting pathological invasiveness

|                        | F1-score | Sensitivity | Specificity | MCC  | ACC  | AUC       |
|------------------------|----------|-------------|-------------|------|------|-----------|
| **Testing set 1**      |          |             |             |      |      |           |
| Deep Learning Model    | 0.545    | 0.600       | 0.900       | 0.464| 0.857| 0.893[0.772,0.100] |
| Radiologists           | 0.554    | 0.800       | 0.767       | 0.481| 0.771| 0.783[0.597,0.969] |
| Thoracic Surgeon (Junior) | 0.422    | 0.367       | 0.932       | 0.340| 0.843| 0.667[0.430,0.904] |
| Thoracic Surgeon (Senior) | 0.367    | 0.300       | 0.950       | 0.360| 0.857| 0.683[0.452,0.914] |
| **Testing set 2**      |          |             |             |      |      |           |
| Frozen Section         | 0.500    | 0.467       | 0.818       | 0.297| 0.708| 0.642[0.397,0.887] |
| Deep Learning Model    | 0.545    | 0.400       | 0.970       | 0.486| 0.857| 0.646[0.429,0.864] |
| Radiologists           | 0.515    | 0.500       | 0.803       | 0.324| 0.708| 0.661[0.422,0.900] |
| Thoracic Surgeon (Junior) | 0.349    | 0.233       | 0.970       | 0.311| 0.740| 0.585[0.324,0.846] |
| Thoracic Surgeon (Senior) | 0.345    | 0.239       | 0.970       | 0.334| 0.740| 0.585[0.324,0.846] |

No significant difference was found between the AUC curves by the DeLong test (P > 0.05).

No significant difference was found between the accuracies by Pearson's χ² test. (P > 0.05)

3.3. Observer study and comparison with the deep learning approach

The accuracies of the two radiologists, two junior thoracic surgeons, and two senior thoracic surgeons ranged from 69.4–83.3% for overall nodules, lower than that of the deep learning model. The mean AUC, sensitivity, specificity, MCC, and F1-score of the six observers are shown in Tables 2 and 3, respectively. For subcentimeter nodules, artificial intelligence models also yielded higher accuracies than six observers (Table 3). Notably, the mean sensitivities of the two radiologists were higher than those of the artificial intelligence models in both testing set 1 and testing sets for overall and subcentimeter nodules.

3.4. Diagnostic accuracy of intraoperative frozen section analysis and comparison with the deep learning approach
The accuracy of frozen sections for overall nodules in testing set 2 was 74.1%, which was lower than that of the deep learning approach (81.0%) (Table 2). Likewise, the accuracy of frozen sections for subcentimeter nodules was 70.8%, lower than the accuracies of the artificial intelligence models (Table 3). Intraoperative frozen section analysis yielded AUC values of 0.755 (95% CI: 66.3–84.7) and 0.642 (95% CI: 39.7–88.7) for overall and subcentimeter nodules, respectively. Compared frozen section analysis, the deep learning approach achieved higher AUC values, at 0.862 (P < 0.05) and 0.646 (P > 0.05), for overall and subcentimeter nodules, respectively (Fig. 3).

4. Discussion

Accurately discriminating pre-IAC from IAC is of great value for preoperative clinical guidance since there are significant differences in the 5-year disease-free survival rate between pre-IAC and IAC (9, 31). AI techniques can capture subtle information from CT images and learn a large number of features or deep representations of a given pulmonary nodule without any additional clinical information. AI techniques integrated with medical images have shown advantages in the invasive classification of lung adenocarcinoma (23, 32, 33). For instance, Wang et al. (21) used 886 ground-glass nodules (GGNs) from 794 patients to predict the invasiveness of lung adenocarcinoma using a deep learning network with an AUC of 0.941. In the clinic, the type of lung adenocarcinoma is identified by histological examination (e.g., biopsy and surgical resection), and diagnosis through CT image review is error prone and time consuming. In our study, the deep learning model achieved good discrimination on both testing set 1 and testing set 2 in terms of the overall nodule size (with AUCs of 0.946 and 0.862, respectively). Although histological examination may still be the gold standard, the method presented in this study provides a convincing, non-invasive method for initial diagnosis before surgical resection.

In this study, the deep learning approach achieved better AUC and accuracy than observers in overall and subcentimeter nodules. The deep learning approach achieved a significantly higher AUC than that of human experts for overall nodules (P < 0.05) in testing set 2. The diagnostic accuracy of well-trained radiologists was slightly lower than that of the deep learning model and higher than the accuracies of thoracic surgeons. Radiologists and surgeons typically focus on visible features such as size, solid components, lesion margin, and other qualitative features, which might be less sensitive to the local evidence that may be exploited by deep learning models. The low accuracy of thoracic surgeons in distinguishing pre-IAC from IAC may relate to the insufficient training and experience of surgeons. Previous studies have reported that deep learning-derived models can achieve equivalent and even higher performance than radiologists; the results of our study support this assertion.

Intraoperative frozen sections are a reliable and routinely used procedure for deciding the extent of surgery (Fig. 2A). This study shows that the deep learning approach achieved comparable performance to frozen sections in determining tumor invasiveness, which could largely improve the current nodule screening process using CT images. For instance, our deep learning model might provide additional information on suspicious nodules, and doctors could integrate this information with patient history and clinical symptoms to guide the treatment plan. Patients with pre-IAC nodules predicted by a deep learning...
model might be more suitable for follow-up monitoring, avoiding invasive surgery. In addition, it only takes a few minutes to detect a patient’s lung nodules in CT images based on AI, while intraoperative frozen sections take hours to complete, which can greatly reduce the patient’s waiting time. Furthermore, to our knowledge, comparisons of the diagnostic accuracy of frozen sections and CT-derived deep learning approaches have not yet been reported. Qiu et al. (34) and Wang et al. (35) compared the diagnostic accuracy of CT-based radiomic methods with that of frozen section analysis for the pathological classification of early-stage lung adenocarcinoma. Qiu et al. (34) reported that the AUC of the nomogram was 0.815, and that of the frozen section analysis was 0.670 ($P = 0.00095$). In this study, the AUC of the deep learning approach was 0.862 in testing set 2 for overall nodules and 0.755 for intraoperative frozen section, which is higher than the study of Qiu et al. (34). The study of Wang et al. (35) classified AAH, AIS, MIA and lepidic predominant adenocarcinoma (LPA) into pre-IAC because of the high 5-year survival of LPA, which made it more difficult for pathologists to distinguish LPA from other invasive adenocarcinomas in frozen sections. This may have contributed to the lower AUC of frozen sections in their study. The study of Wang et al. (35) reported no significant difference in the overall diagnostic accuracy between the radiomics method and FS (68.8% vs. 70.0%, $P = 0.836$), which is consistent with the results of our study.

Clinically, many factors affect intraoperative frozen section diagnoses, such as tumor size, sampling issues, and even nodule density. Liu et al. (36) reported that the diagnostic accuracy of FS for tumors smaller than 1 cm and larger than 1 cm in diameter was 79.6% and 90.8%, respectively. Yeh et al. (37) reported an average frozen section diagnostic accuracy of 64% (54–74%) for discriminating among AIS, MIA, and invasive adenocarcinomas by five pathologists. In this study, the accuracies of frozen sections for overall and subcentimeter nodules were 74.1% and 70.8%, respectively. Discrepancies were mostly due to the underestimation of AIS and MIA. A high percentage of AIS/MIA and concurrent subcentimeter nodules may be one of the reasons for the high accuracy of the study of Liu et al. (36). Moreover, Zhu et al. (38) analyzed 803 cases and reported that misdiagnosis by frozen sections because of sampling error might lead to incomplete resection. Our study results suggest that a deep learning approach could serve as a reliable and complementary method when pathological evaluation cannot be performed intraoperatively.

However, this study still has several limitations. First, this is a retrospective study conducted at a single institution and is therefore subject to potential biases concerning patient selection, measurements, and observers. Prospective and multicenter trials are required in future studies. Second, intraoperative frozen sections also aid in determining the resection margin, which is not supported yet in the proposed deep learning approach. Therefore, the another interesting research direction for the deep learning approach is to estimate appropriate surgery margin in clinical application. Third, efficient integration of the deep learning approach into clinical workflows still needs to be explored. Fourth, sample size of subcentimeter nodules in the testing set was relatively low, which may decrease the model generalizability. Future work should include a large number of subcentimeter nodules to improve the performance of a deep learning approach in predicting tumor invasiveness.
5. Conclusion

We established a deep learning approach that demonstrated plausible performance, and its ability to
distinguish tumor invasiveness was comparable to that of intraoperative frozen section analysis. This
deep learning approach has potential value in clinically guiding surgical strategies, but it still needs to be
verified in prospective and multicenter trials.

Abbreviations

AUC, area under the curve; NSCLC, non-small cell lung cancer; IASLC, International Association for the
Study of Lung Cancer; ATS, American Thoracic Society; ERS, European Respiratory Society; AIS,
adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive lung adenocarcinoma;
pre-IAC, preinvasive lung adenocarcinoma; AI, artificial intelligence; CNN, convolutional neural network;
LPA, lepidic predominant adenocarcinoma.

Declarations

Ethical statement

The authors are responsible 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 was approved by the
Institutional Review Board of Shanghai Chest Hospital, and individual consent for this retrospective
analysis was waived.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on
reasonable request.

Conflicts of interest

The authors have no conflicts of interest to declare.

Funding

This work was supported by the Interdisciplinary Program of Shanghai Jiao Tong University (grant no.
YG2014QN22), Cooperative Research Project of Shanghai Jiao Tong University Collaborative Innovation
Center of Translational Medicine (TM201822) and Shanghai Science and Technology Support Project
(No.).
Author contributions

(I) Conception and design: Yilv Lv, Yiqiang Zhan, Bo Ye;

(II) Administrative support: Yiqiang Zhan, Yuchen Han, Hong Yu;

(III) Provision of study materials or patients: Bo Ye, Yuchen Han, Hong Yu;

(IV) Collection and assembly of data: Yilv Lv, Rong Hua, Jia Huang, Kuan Xu, Yaozong Gao, Xianjie Zhang, Yifan Wu;

(V) Data analysis and interpretation: Yilv Lv, Ying Wei, Xiaobin Zhang, Min Li, Cui Tang, Long Yang, Bingchun Liu, Yuchen Han, Zhanxian Shang;

(VI) Manuscript writing: All authors

(VII) Final approval of manuscript: All authors

Acknowledgements

We thank Dr. Zhichao Liu for providing valuable suggestions in the revision of this manuscript.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-49.

2. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2021. CA Cancer J Clin 2021;71:7-33.

3. Yang D, Liu Y, Bai C, et al. Epidemiology of lung cancer and lung cancer screening programs in China and the United States. Cancer Lett 2020;468:82-7.

4. 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 22-3.

5. Landreneau RJ, Normolle DP, Christie NA, et al. Recurrence and survival outcomes after anatomic segmentectomy versus lobectomy for clinical stage I non-small-cell lung cancer: a propensity-matched analysis. J Clin Oncol 2014;32:2449-55.

6. El-Sherif A, Gooding WE, Santos R, et al. Outcomes of sublobar resection versus lobectomy for stage I non-small cell lung cancer: a 13-year analysis. Ann Thorac Surg 2006;82:408-15; discussion 15-6.

7. Suzuki K, Watanabe SI, Wakabayashi M, et al. A single-arm study of sublobar resection for ground-glass opacity dominant peripheral lung cancer. J Thorac Cardiovasc Surg 2020. doi: 10.1016/j.jtcvs.2020.09.146.

8. 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.
9. Yanagawa N, Shiono S, Abiko M, et al. New IASLC/ATS/ERS classification and invasive tumor size are predictive of disease recurrence in stage I lung adenocarcinoma. J Thorac Oncol 2013;8:612-8.
10. Huang KY, Ko PZ, Yao CW, et al. Inaccuracy of lung adenocarcinoma subtyping using preoperative biopsy specimens. J Thorac Cardiovasc Surg 2017;154:332-9.e1.
11. Huang CS, Hsu PK, Chen CK, et al. Preoperative biopsy and tumor recurrence of stage I adenocarcinoma of the lung. Surg Today 2019;50:673-84.
12. Ahn SY, Yoon SH, Yang BR, et al. Risk of pleural recurrence after percutaneous transthoracic needle biopsy in stage I non-small-cell lung cancer. Eur Radiol 2019;29:270-8.
13. Inoue M, Honda O, Tomiyama N, et al. Risk of pleural recurrence after computed tomographic-guided percutaneous needle biopsy in stage I lung cancer patients. Ann Thorac Surg 2011;91:1066-71.
14. 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.
15. 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.
16. Lee SM, Park CM, Goo JM, et al. Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features. Radiology 2013;268:265-73.
17. Cohen JG, Reymond E, Lederlin M, et al. Differentiating pre- and minimally invasive from invasive adenocarcinoma using CT-features in persistent pulmonary part-solid nodules in Caucasian patients. Eur J Radiol 2015;84:473-88.
18. Zhang Y, Shen Y, Qiang JW, et al. HRCT features distinguishing pre-invasive from invasive pulmonary adenocarcinomas appearing as ground-glass nodules. Eur Radiol 2016;26:2921-8.
19. Niu R, Shao X, Shao X, et al. Lung adenocarcinoma manifesting as ground-glass opacity nodules 3 cm or smaller: evaluation with combined high-resolution CT and PET/CT modality. AJR Am J Roentgenol 2019;213:W236-45.
20. Zhan Y, Peng X, Shan F, et al. Attenuation and morphologic characteristics distinguishing a ground-glass nodule measuring 5-10 mm in diameter as invasive lung adenocarcinoma on thin-slice CT. AJR Am J Roentgenol 2019;213:W162-70.
21. Wang X, Li Q, Cai J, et al. Predicting the invasiveness of lung adenocarcinomas appearing as ground-glass nodule on CT scan using multi-task learning and deep radiomics. Transl Lung Cancer Res 2020;9:1397-406.
22. Xia X, Gong J, Hao W, et al. Comparison and fusion of deep learning and radiomics features of ground-glass nodules to predict the invasiveness risk of stage-I lung adenocarcinomas in CT scan. Front Oncol 2020;10:418.
23. Zhao W, Yang J, Sun Y, et al. 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. Cancer Res 2018;78:6881-9.
24. Park S, Park G, Lee SM, et al. Deep learning-based differentiation of invasive adenocarcinomas from preinvasive or minimally invasive lesions among pulmonary subsolid nodules. Eur Radiol 2021;31:6239-47.

25. Chaunzwa TL, Hosny A, Xu Y, et al. Deep learning classification of lung cancer histology using CT images. Sci Rep 2021;11:5471.

26. Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436.

27. Ashraf SF, Yin K, Meng CX, et al. Predicting benign, preinvasive, and invasive lung nodules on computed tomography scans using machine learning. J Thorac Cardiovasc Surg 2021. doi: 10.1016/j.jtcvs.2021.02.010.

28. Gu D, Liu G, Xue Z. On the performance of lung nodule detection, segmentation and classification. Comput Med Imaging Graph 2021;89:101886.

29. Detterbeck FC, Boffa DJ, Kim AW, et al. The eighth edition lung cancer stage classification. Chest 2017;151:193-203.

30. Mu G, Chen Y, Wu D, et al. Relu cascade of feature pyramid networks for CT pulmonary nodule detection. In: Suk H-I, Liu M, Yan P, Lian C, eds. Machine Learning in Medical Imaging. Cham: Springer International Publishing; 2019. pp. 444-52.

31. You J, Zhang W, Zhang Q, et al. Does lung adenocarcinoma subtype predict patient survival?: a clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenoc. J Thorac Oncol 2011;6:1496-504.

32. Sun Y, Li C, Jin L, et al. Radiomics for lung adenocarcinoma manifesting as pure ground-glass nodules: invasive prediction. Eur Radiol 2020;30:3650-9.

33. Gong J, Liu J, Hao W, et al. A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images. Eur Radiol 2020;30:1847-55.

34. Qiu ZB, Zhang C, Chu XP, et al. Quantifying invasiveness of clinical stage IA lung adenocarcinoma with computed tomography texture features. J Thorac Cardiovasc Surg 2020. doi: 10.1016/j.jtcvs.2020.12.092.

35. 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.

36. 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.

37. 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.

38. 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-
**Figures**

**Figure 1**

The flow chart of patient selection and deep learning architecture. “Conv” represents a convolution, “k” represents the kernel, and “s” denotes the number of strides. “BN” represents the batch normalization layer. “FC” represents a fully connected layer.
Figure 2

Diagram of current and artificial intelligence procedures to determine histological invasiveness.
Figure 3

ROC curves showing the performance of the deep learning model and current methods in distinguishing pre-IAC from IAC in the independent testing set. A, ROC curve for overall nodules. The AUCs of the deep learning model were 0.862 and 0.755 for frozen sections. The difference between the two methods was statistically significant (P=0.043). B, ROC curve for subcentimeter nodules. The AUC was 0.646 for the deep learning model and 0.642 for frozen sections (P=0.976).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryFig.1.tif
- supplementarytable1.docx