Journal of the National Cancer Center 4 (2024) 233–240

Contents lists available at ScienceDirect

Journal of the National Cancer Center

journal homepage: www.elsevier.com/locate/jncc

Full Length Article

Deep learning model based on primary tumor to predict lymph node status in clinical stage IA lung adenocarcinoma: a multicenter study

Li Zhang[1,†], Hailin Li[2,6,†], Shaohong Zhao[3,†], Xuemin Tao[3], Meng Li[1], Shouxin Yang[4], Lina Zhou[1], Mengwen Liu[1], Xue Zhang[1], Di Dong[5,6,*], Jie Tian[2,6,*], Ning Wu[1,*]

[1] *Department of Diagnostic Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China*
[2] *Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, China*
[3] *Department of Radiology, PLA General Hospital, Beijing, China*
[4] *Department of Radiology, Peking University Cancer Hospital & Institute, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Beijing, China*
[5] *School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China*
[6] *CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

Keywords:
Lung neoplasm
Adenocarcinoma
Clinical stage IA
Deep learning
Lymph node status

ABSTRACT

Objective: To develop a deep learning model to predict lymph node (LN) status in clinical stage IA lung adenocarcinoma patients.

Methods: This diagnostic study included 1,009 patients with pathologically confirmed clinical stage T1N0M0 lung adenocarcinoma from two independent datasets (699 from Cancer Hospital of Chinese Academy of Medical Sciences and 310 from PLA General Hospital) between January 2005 and December 2019. The Cancer Hospital dataset was randomly split into a training cohort (559 patients) and a validation cohort (140 patients) to train and tune a deep learning model based on a deep residual network (ResNet). The PLA Hospital dataset was used as a testing cohort to evaluate the generalization ability of the model. Thoracic radiologists manually segmented tumors and interpreted high-resolution computed tomography (HRCT) features for the model. The predictive performance was assessed by area under the curves (AUCs), accuracy, precision, recall, and F1 score. Subgroup analysis was performed to evaluate the potential bias of the study population.

Results: A total of 1,009 patients were included in this study; 409 (40.5%) were male and 600 (59.5%) were female. The median age was 57.0 years (inter-quartile range, IQR: 50.0–64.0). The deep learning model achieved AUCs of 0.906 (95% CI: 0.873–0.938) and 0.893 (95% CI: 0.857–0.930) for predicting pN0 disease in the testing cohort and a non-pure ground glass nodule (non-pGGN) testing cohort, respectively. No significant difference was detected between the testing cohort and the non-pGGN testing cohort (P = 0.622). The precisions of this model for predicting pN0 disease were 0.979 (95% CI: 0.963–0.995) and 0.983 (95% CI: 0.967–0.998) in the testing cohort and the non-pGGN testing cohort, respectively. The deep learning model achieved AUCs of 0.848 (95% CI: 0.798–0.898) and 0.831 (95% CI: 0.776–0.887) for predicting pN2 disease in the testing cohort and the non-pGGN testing cohort, respectively. No significant difference was detected between the testing cohort and the non-pGGN testing cohort (P = 0.657). The recalls of this model for predicting pN2 disease were 0.903 (95% CI: 0.870–0.936) and 0.931 (95% CI: 0.901–0.961) in the testing cohort and the non-pGGN testing cohort, respectively.

Conclusions: The superior performance of the deep learning model will help to target the extension of lymph node dissection and reduce the ineffective lymph node dissection in early-stage lung adenocarcinoma patients.

* Corresponding authors.
E-mail addresses: di.dong@ia.ac.cn (D. Dong), jie.tian@ia.ac.cn (J. Tian), cjr.wuning@vip.163.com (N. Wu).
† These authors contributed equally to this work.

https://doi.org/10.1016/j.jncc.2024.01.005
Received 15 September 2023; Received in revised form 28 December 2023; Accepted 22 January 2024

1. Introduction

Lung cancer is the leading cause of cancer incidence and mortality worldwide, and non-small cell lung cancer (NSCLC) accounts for 80% of the lung cancer cases.[1] Adenocarcinoma is the most common subtype of NSCLC. Increasing peripheral small lung adenocarcinomas are detected due to the advent of high-resolution computed tomography (HRCT) and chest low-dose computed tomography (CT) screening.[2,3] It was reported that about 20% of lung adenocarcinoma patients had lymph node (LN) metastasis after systematic LN dissection, even though they were confirmed as clinical N0 before surgical resection.[4] Therefore, systematic LN dissection is still the preferred treatment for these patients to avoid tumor residuals and prolong survival. However, Rami-porta et al. studied the pathological staging of 77,156 cases of lung cancer diagnosed from 1999 to 2010 recruited by the International Association for the Study of Lung Cancer (IASLC), and found that about 25% of lung adenocarcinoma were in stage IA at the time of diagnosis.[5] This result means that about 1/4 of patients with lung adenocarcinoma have undergone ineffective LN dissection, which may lead to unnecessary tissue damage, operational risk, and longer postoperative recovery time.[6,7] Therefore, accurate prediction of lymph node metastasis is crucial for surgeons in determining the necessity of lymph node dissection and selecting the specific node stations for the procedure. This approach helps avoid unnecessary interventions while ensuring adequate treatment for cases with higher risk of metastasis.

Researchers have been trying to find a method or some risk factors to accurately evaluate the status of LN in clinical N0 patients. Articles focusing on morphological features, positron emission tomography (PET) metabolic index, and texture characteristics of primary tumor or serum markers such as carcinoembryonic antigen have been published in recent years.[4,8,9] However, the ability of these features to preoperatively evaluate lymph nodes status is not ideal, with the area under the curve (AUC) being less than 0.8, which limits their promotion and application in clinical practice. Radiomics, which can convert medical radiographic images into high-dimensional, mineable, and quantitative data, shows its strong ability in oncology, including the diagnosis of LN metastasis.[10-15] The reported AUC of radiomics signatures for predicting LN metastasis in lung cancer patients can achieve more than 0.80.[10,16,17] However, radiomics models are very sensitive to image scanning parameters, and different centers often get different results, which limits the clinical application of the models.[18] Moreover, predefined features based on medical images sometimes experience problems. For example, inter-group or intra-group differences in manual segmentation of lesions or high false positives in feature selection can limit the robustness and generalization of traditional machine learning models.

End-to-end deep learning algorithms, however, can automatically learn the potential connections between images and labels without extracting pre-defined image features.[19-22] Moreover, neural network models have significantly advanced beyond traditional machine learning methods in terms of both model construction and application. One notable example of this advancement is the deep residual networks, commonly known as ResNet, which is particularly notable for its flexible structure and exceptional capabilities in image classification tasks. This flexibility and efficiency have made it a popular choice in the field of deep learning, especially for tasks that involve complex image processing and recognition.[23] Furthermore, evidence suggests that integrating information from diverse modalities, including images and clinical features, enhances the diagnostic precision of artificial intelligence models.[24-26] In this study, we designed a deep learning model based on ResNet and combined CT and clinical information to predict LN status before surgery for stage IA lung adenocarcinoma patients to help target the extension of LN dissection and reduce the ineffective LN dissection in early-stage lung adenocarcinoma patients.

2. Methods

2.1. Case selection and evaluation of clinical characteristics

A total of 1,009 pathologically confirmed clinical stage T1N0M0 lung adenocarcinoma patients between January 2005 and December 2019 from two independent datasets were enrolled in this study. Data analysis was performed between January 2022 and March 2023. These patients included 699 cases from the Cancer Hospital of Chinese Academy of Medical Sciences (Cancer Hospital) and 310 cases from PLA General Hospital (PLA Hospital). The inclusive criteria were: (a) pathologically proven lung adenocarcinoma patients; (b) systematic LN dissection performed; and (c) contrast-enhanced thin-section CT performed less than two weeks before surgical resection. The exclusive criteria were: (a) multiple lesions or evidence of metastasis; (b) tumor diameter of more than 3 cm; (c) short diameter of the LN in hila or mediastina larger than 1 cm on HRCT image; (d) preoperative therapy (radiotherapy, chemotherapy, or chemoradiotherapy); and (e) malignancy history in recent five years.

Clinicopathological characteristics were recorded for all eligible patients. All pathological sections were interpreted according to the multidisciplinary adenocarcinoma criteria.[27] Resected specimens of patients in 2005–2012 were reviewed by an experienced lung pathologist (S.W., with 8 years of experience in lung pathology) according to the same criteria. Atypical adenomatous hyperplasia (AAH) and adenocarcinoma in situ (AIS) were integrated to precursors of adenocarcinoma (PGL) according to the 5th World Health Organization (WHO) classification of lung tumors.[28] We performed tumor staging according to the 8th Edition of the American Joint Committee on Cancer TNM Staging Manual.[29]

Two independent thoracic radiologists (L.Z. & M.L., with more than 15 years of experience) reviewed the preoperative CT images on GE AW 4.6 workstation blinded to pathological results. When there was a discrepancy in the interpretation of morphological features, a final consensus was reached by group discussions. 14 features of emphysema, bronchiectasis, invasive lobe, location, nodule consistency, diameter, shape, boundary, enhancement, pleural retraction, deep lobulation, calcification, necrosis, and regional air spaces were interpreted. The definitions of these features were listed in Supplementary Table 1 in Supplementary materials.

2.2. Development of deep learning model

The Cancer Hospital dataset was divided into a training cohort and a validation cohort at a ratio of 8:2 to train and tune the deep learning model (559 patients for the training cohort and 140 patients for the validation cohort. CT slices from any individual patient were exclusively assigned to one cohort), and patients from the PLA Hospital were used as a testing cohort to verify the generalization ability of the deep learning model. The model training was performed on the training cohort, the hyper-parameter optimization and epoch selection were performed on the validation cohort, and the testing cohort was only used to evaluate the model performance.

To obtain enough data to fully fit the deep learning model, we replaced the 3D CT volume with 2D CT slices and processed each CT slice independently. An experienced radiologist with 15 years of practice, Dr. Z.L., meticulously delineated the tumor region on each CT slice for all 1,009 patients, using ITK-SNAP software (version 3.6.0; www.itksnap.org). Subsequently, to focus the deep learning model's attention on the tumor area and its adjacent peritumoral microenvironment, we expanded the region of interest (ROI) of the primary tumor to include the surrounding lung parenchyma, enlarging it threefold. This led to the generation of axis-aligned bounding boxes from the manual segmentations on each slice. For converting these bounding boxes into square images, we adapted the shorter sides of the rectan-

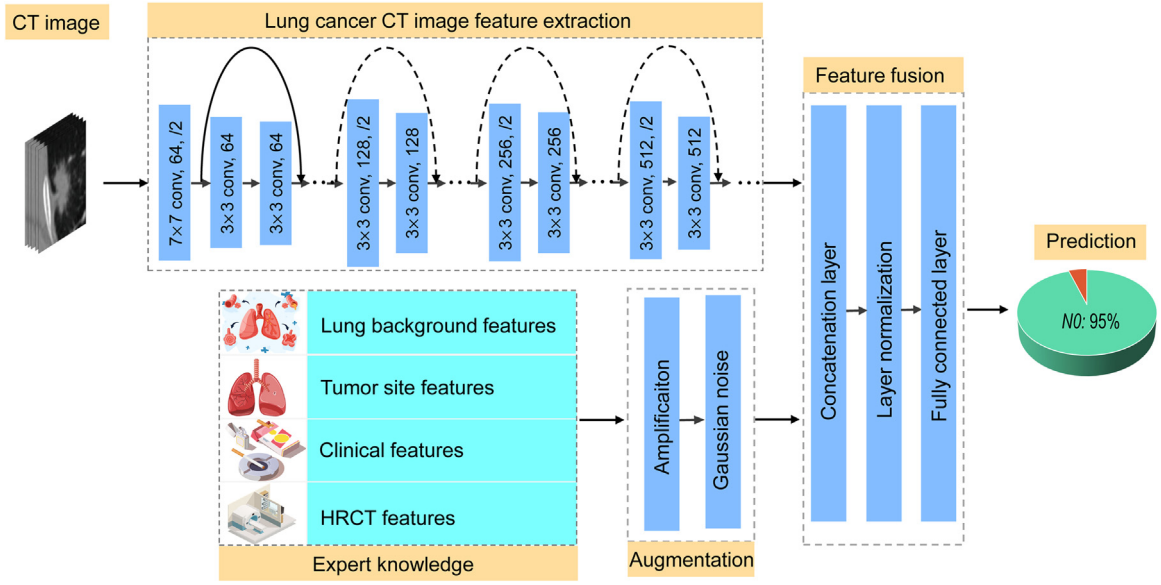

Fig. 1. The architecture of the deep-learning model. CT, computed tomography; HRCT, high-resolution computed tomography.

gular bounding boxes, extending them equally on both sides to match the lengths of the longer sides. These uniformly resized square images then served as the inputs for the deep learning model. To maintain quality, we excluded any bounding boxes smaller than 64×64 pixels, ensuring that all input images were of adequate size and resolution for effective processing. Finally, 13,073 slices of primary tumor CT images were generated from 1,009 patients to train, validate, and test the deep learning model. Of these 13,073 slices of primary tumor CT images, 9,569 (73.2%) were pN0 disease, 2,004 (15.3%) were pN1 disease and 1,500 (11.5%) were pN2 disease. 7,811 slices were assigned to the training cohort (5,623 [72.0%] were pN0 disease, 1,217 [15.6%] were pN1 disease, and 971 [12.4%] were pN2 disease), 2,013 slices were assigned to the validation cohort (1,480 [73.5%] were pN0 disease, 395 [19.6%] were pN1 disease, and 138 [6.9%] were pN2 disease), and 3,249 slices were assigned to the testing cohort (2,466 [75.9%] were pN0 disease, 392 [12.1%] were pN1 disease, and 391 [12.0%] were pN2 disease).

We also incorporated 17 clinic-radiological characteristics into the model so that the model could learn information from different modalities. Of these 17 features, three were clinical features including sex, age, and smoking history, two were lung background features including emphysema and bronchiectasis, two were tumor sites including invasive lobe and location, and ten were CT features including nodule consistency, diameter, shape, boundary, enhancement, pleural retraction, deep lobulation, calcification, necrosis, and regional air spaces.

Then, we designed an end-to-end deep learning model based on ResNet-34. The model combined the two modalities by concatenating the last feature map of the ResNet architecture with clinic-radiological features. Given that the number of feature maps significantly outnumbered the clinic-radiological features, we enhanced the 17 clinic-radiological characteristics to a 128-dimensional set through replication, thus preventing information loss. Specifically, we expanded the initial set of 17 clinical features by replicating each feature seven times, and augmented this set with 9 randomly selected values, culminating in a 128-dimensional clinical feature set. Subsequently, these 128-dimensional clinic-radiological features were concatenated with the 512-dimensional feature map, which was automatically extracted by ResNet34. Finally, the two types of features were normalized by layer normalization and then fed to a fully connected layer. The model architecture is shown in Fig. 1.

During the model training, cropping, and flipping were used to augment the input image. Given the difference in the number of CT slices between patients with LN metastasis and those without, we adopted weighted random sampling to balance the number of CT slices in each training iteration. To ensure the robustness of our model's training process, we strategically addressed the potential correlation among slices from the same patient by avoiding the inclusion of neighboring slices in the same training batch. We used cross-entropy loss as the loss function and Adam as the optimization algorithm. The model was trained for 100 epochs, and the learning rate decreased from 1e-3 to 1e-5 by stepwise decay at every 5 epochs.

2.3. Assessment of deep learning model

After the deep learning model prediction of each CT slice was completed, we integrated the prediction results of slices from the same patient to calculate a patient-level prediction result. A voting-based method was used to map the prediction results at the slice level to the patient level. The formula is as follows:

$$P_i = f\left(\sum_{j=1}^{n_i} \frac{sgn\left(s_{i,j} - 0.5\right) + 1}{2n_i}\right),$$

$$p_i = \frac{\sum_j s_{i,j}}{card\left(\mathcal{N}\right)} \qquad j \in \mathcal{N},$$

where

$$f(x) = \frac{sgn(x - 0.5) + 1}{2},$$

$$\mathcal{N} = \left\{j \mid f\left(s_{i,j}\right) = P_i\right\}$$

n_i: The total number of slices for the *i-th* patient.

P_i: The predictive value of the *i-th* patient. 0 represents negative and 1 represents positive.

p_i: The risk probability of LN metastasis for the *i-th* patient.

$s_{i,j}$: The probability that the *j-th* slice of the *i-th* patient has positive predictions.

Then, receiver operating characteristic (ROC) curve analysis was performed to evaluate the performance of the deep learning model, and AUCs were calculated to quantify the performance. The accuracy, precision, recall, and F1 score results were also calculated. Delong test was

Table 1
Clinicopathological characteristics of clinical T1N0M0 patients.

	Training cohort (n = 559)	Validation cohort (n = 140)	Testing cohort (n = 310)
Data source	Cancer Hospital	Cancer Hospital	LPA Hospital
Age, median (IQR), years	58.0 (50.0–64.0)	58.0 (49.0–64.0)	56.0 (51.0–63.0)
Sex, n (%)			
Female	350 (62.6)	84 (60.0)	166 (53.5)
Male	209 (37.4)	56 (40.0)	144 (46.5)
Smoking history, n (%)			
Smoker	137 (24.5)	38 (27.1)	94 (30.3)
Non-smoker	422 (75.5)	102 (72.9)	216 (69.7)
Surgical procedure, n (%)			
Wedge resection	50 (8.9)	4 (2.9)	9 (2.9)
Segmentectomy	35 (6.3)	7 (5.0)	16 (5.2)
Lobectomy	474 (84.8)	129 (92.1)	285 (91.9)
Tumor location, n (%)			
Left	199 (35.6)	63 (45.0)	108 (34.8)
Right	360 (64.4)	77 (55.0)	202 (65.2)
Pathological N stage, n (%)			
N0	440 (78.7)	112 (80.0)	244 (78.7)
N1	68 (12.2)	20 (14.3)	35 (11.3)
N2	51 (9.1)	8 (5.7)	31 (10.0)
Pathological stage, n (%)			
0 /Tis	61 (10.9)	10 (7.1)	4 (1.30)
A	310 (55.5)	82 (58.6)	210 (67.7)
B	69 (12.3)	20 (14.3)	30 (9.7)
B	68 (12.2)	20 (14.3)	35 (11.3)
A	51 (9.1)	8 (5.7)	31 (10.0)
Histological subtype, n (%)			
Precursors of adenocarcinoma	62 (11.1)	10 (7.1)	4 (1.3)
MIA	51 (9.1)	13 (9.3)	28 (9.0)
IAC	446 (79.8)	117 (83.6)	278 (89.7)

Abbreviations: IAC, invasive adenocarcinoma; IQR, inter-quartile range; MIA, minimally invasive adenocarcinoma; N, node; Tis, tumor in situ.

used to compare the model performance in the testing cohort between all patients and non-pGGN patients.

2.4. Statistical analysis

In this study, the model construction was implemented using Pytorch (version 1.7.1, http://www.pytorch.org/) in Python (version 3.8.5, https://www.python.org/). The performance evaluation was conducted in R (version 3.5.1, https://www.r-project.org/). The levels of statistical significance of the reports were all two-sided, and P values < 0.05 were considered to indicate significance. The detailed packages used in Python and R were described in Supplementary Table 2.

3. Results

3.1. Clinicopathological characteristics of patients

Of the 1,009 clinical stage T1N0M0 adenocarcinoma patients, 409 (40.5%) were male and 600 (59.5%) were female. The median age was 57.0 years (inter-quartile range, IQR: 50.0–64.0), 740 (73.3%) were non-smokers, and 269 (26.7%) were smokers. 63 (6.3%) underwent wedge resection, 58 (5.7%) underwent segmentectomy, and 888 (88.0%) underwent lobectomy. The primary tumors of 370 (36.7%) patients were located in the left lung and those of 639 (63.3%) patients were in the right lung.

Lymph node status were negative in 796 (78.9%) of the total clinical stage T1N0M0 adenocarcinoma patients, and positive in 213 (21.1%) of them, among whom 123 patients were with N1 disease, and 90 patients were with N2 disease. The pathological stages were stage 0 (Tis) in 75 (7.4%), stage IA in 602 (59.7%), stage IB in 119 (11.8%), stage IIB in 123 (12.2%), and stage IIIA in 90 (8.9%) patients. Histological subtypes were precursors of adenocarcinoma in 76 (7.5%), minimally invasive adenocarcinoma (MIA) in 92 (9.1%), and invasive adenocarcinoma (IAC) in 841 (83.4%) patients. The clinicopathological characteristics of patients in the training cohort, the validation cohort, and the testing cohort are shown in Table 1.

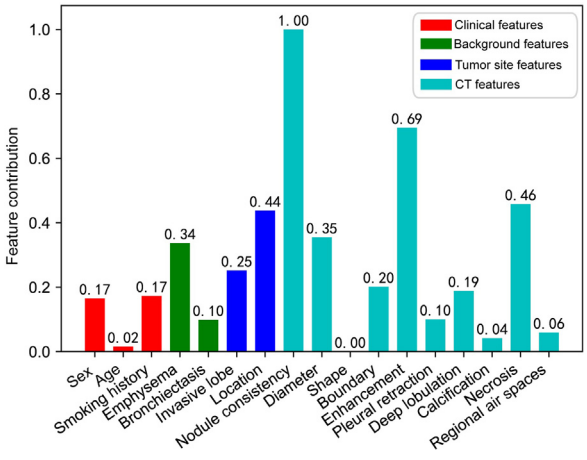

Fig. 2. The contributions of clinic-radiological features to the deep learning model. CT, computed tomography.

3.2. Contributions of clinic-radiological features to the deep learning model

The contributions of clinic-radiological features to the deep learning model were shown in Fig. 2. Nodule consistency contributes the most, followed by enhancement, necrosis, location, diameter, emphysema, invasive lobe, boundary, deep lobulation, sex, smoking history, bronchiectasis, pleural retraction, regional air space, calcification, age, and shape. The clinic-radiological features in the training cohort, the validation cohort, and the testing cohort were shown in Supplementary Table 3.

Table 2
The performance of the deep learning model in predicting pathological N0 and N2 diseases.

	AUC (95% CI)	Accuracy (95% CI)	F1 Score (95% CI)	Precision (95% CI)	Recall (95% CI)
All patients					
N0 disease					
Training	0.963 (0.949–0.978)	0.903 (0.879–0.928)	0.935 (0.914–0.955)	1.000 (1.000–1.000)	0.877 (0.850–0.904)
Validation	0.932 (0.893–0.971)	0.857 (0.799–0.915)	0.906 (0.857–0.954)	0.960 (0.928–0.993)	0.857 (0.799–0.915)
Testing	0.906 (0.873–0.938)	0.800 (0.755–0.845)	0.857 (0.818–0.896)	0.979 (0.963–0.995)	0.762 (0.715–0.810)
N2 disease					
Training	0.918 (0.894–0.942)	0.778 (0.744–0.813)	0.451 (0.410–0.493)	0.291 (0.254–0.329)	1.000 (1.000–1.000)
Validation	0.905 (0.815–0.996)	0.757 (0.686–0.828)	0.292 (0.216–0.367)	0.175 (0.112–0.238)	0.875 (0.820–0.930)
Testing	0.848 (0.798–0.898)	0.690 (0.639–0.74)	0.368 (0.315–0.422)	0.231 (0.184–0.278)	0.903 (0.870–0.936)
Non-pGGN patients					
N0 disease					
Training	0.952 (0.933–0.970)	0.881 (0.851–0.911)	0.912 (0.886–0.938)	1.000 (1.000–1.000)	0.839 (0.805–0.873)
Validation	0.915 (0.866–0.963)	0.829 (0.761–0.897)	0.880 (0.821–0.939)	0.948 (0.908–0.988)	0.820 (0.751–0.890)
Testing	0.893 (0.857–0.930)	0.805 (0.759–0.851)	0.861 (0.820–0.901)	0.983 (0.967–0.998)	0.766 (0.716–0.815)
N2 disease					
Training	0.897 (0.866–0.927)	0.727 (0.686–0.768)	0.451 (0.406–0.497)	0.291 (0.250–0.333)	1.000 (1.000–1.000)
Validation	0.885 (0.776–0.994)	0.709 (0.627–0.792)	0.292 (0.209–0.374)	0.175 (0.106–0.244)	0.875 (0.815–0.935)
Testing	0.831 (0.776–0.887)	0.699 (0.645–0.752)	0.388 (0.332–0.445)	0.245 (0.195–0.296)	0.931 (0.901–0.961)

Abbreviations: AUC, area under curve; CI, confidence interval; pGGN, pure ground glass nodule.

To further assess the impact of integrating clinic-radiological features, we developed a baseline neural network model devoid of clinic-radiological feature fusion. This model was based on the same backbone (ResNet34) and subjected to the same dataset conditions for a comparative analysis. The findings distinctly demonstrate that our proposed fusion deep learning model significantly surpasses traditional single-modality deep learning models across a variety of tasks and subgroups. Detailed results can be found in Supplementary Table 4.

3.3. Performance of deep learning model on predicting lymph node status of all patients

The accuracy at the slice level on the training cohort, the validation cohort, and the testing cohort was 0.904, 0.826, and 0.793, respectively. The sensitivity and specificity were relatively balanced in the three cohorts (sensitivity: 0.952, 0.869, 0.890 for the training, validation, and testing cohorts, respectively; specificity: 0.846, 0.811, 0.762 for the training, validation, and testing cohorts, respectively).

At the patient level, the deep learning model reached an AUC of 0.906 (95% CI: 0.873–0.938) in the testing cohort for predicting pathological N0 disease. The model achieved a high F1 score of 0.857 (95% CI: 0.818–0.896). The accuracy, precision, and recall of the model for predicting pathological N0 disease was 0.800 (95% CI: 0.755–0.845), 0.979 (95% CI: 0.963–0.995), and 0.762 (95% CI: 0.715–0.810), respectively. For predicting N2 disease, the AUC in the testing cohort was 0.848 (95% CI: 0.798–0.898), with an F1 score of 0.368 (95% CI: 0.315–0.422). The accuracy, precision, and recall of the model for predicting pathological N2 disease was 0.690 (95% CI: 0.639–0.740), 0.231 (95% CI: 0.184–0.278), and 0.903 (95% CI: 0.870–0.936), respectively. The performances of the model in the training and validation cohorts were demonstrated in Table 2.

3.4. Performance of deep learning model on predicting lymph node status of non-pure ground glass nodule (non-pGGN) patients

Since there were very few LN metastases in pGGN patients, we also performed ROC curve analysis for the non-pGGN patients in the three cohorts. There were 853 non-pGGN patients in total, consisting of 402 patients (47.1%) in the training cohort, 169 patients (19.8%) in the validation cohort, and 282 patients (33.1%) in the testing cohort. The pathological N stages for these 402 non-pGGN patients in the training cohort were N0 in 303 (75.4%), N1 in 54 (13.4%), and N2 in 45 (11.2%) patients. The pathological N stages for these 169 non-pGGN patients in the validation cohort were N0 in 121 (71.6%), N1 in 34 (20.1%), and N2 in 14 (8.3%) patients. The pathological N stages for these 282 non-pGGN patients in the testing cohort were N0 in 222 (78.7%), N1 in 31 (11.0%), and N2 in 29 (10.3%) patients.

The deep learning model reached an AUC of 0.893 (95% CI: 0.857–0.930) in the non-pGGN testing cohort for predicting pathological N0 disease, which was lower than that in the whole testing cohort, but no significant difference was detected ($P = 0.622$). The deep learning model reached an AUC of 0.831 (95% CI: 0.776–0.887) in the non-pGGN testing cohort for predicting pathological N2 disease, which was lower than that in the whole testing cohort, but no significant difference was detected ($P = 0.657$). The precision of the model for predicting pN0 disease in the non-pGGN testing cohort was 0.983 (95% CI: 0.967–0.998). The recall of the model for predicting pN2 disease in the non-pGGN testing cohort was 0.931 (95% CI: 0.901–0.961). The results of ROC analysis were shown in Fig. 3. The detailed performance was shown in Table 2.

4. Discussion

Reliable predictive methods for LN status are crucial to optimize surgical schemes in early-stage lung adenocarcinoma patients. Our findings suggested that the deep learning model based on the primary tumor image data and clinic-radiological data had a superior performance on pathological LN status prediction in clinical T1N0M0 lung adenocarcinoma patients. The AUCs of the model for predicting pathological N0 disease reached 0.906 (95% CI: 0.873–0.938) in the testing cohort and 0.893 (95% CI: 0.857–0.930) in non-pGGN patients of the testing cohort, with precision of 0.979 and 0.983, respectively, while the AUC of the model for predicting N2 disease was 0.848 (95% CI: 0.798–0.898) in the testing cohort and 0.831 (95% CI: 0.776–0.887) in the non-pGGN patients of the testing cohort, with a recall of 0.903 and 0.931, respectively. No significant difference in the performance of the model was detected between the whole testing cohort and the non-pGGN testing cohort ($P = 0.622$ for pN0 disease and $P = 0.657$ for pN2 disease).

The most used non-invasive preoperatively LN evaluation methods for lung cancer patients are chest CT and positron emission tomography (PET)-CT at present. When taking axial short diameter less than 1 cm on CT image and SUVmax less than 2.5 on PET-CT as diagnostic criteria, the precision is about 0.800 for predicting pN0 disease.[30-32] In this study, the precision of the deep learning model for predicting pN0 disease was massively improved (0.979 and 0.983 in the testing

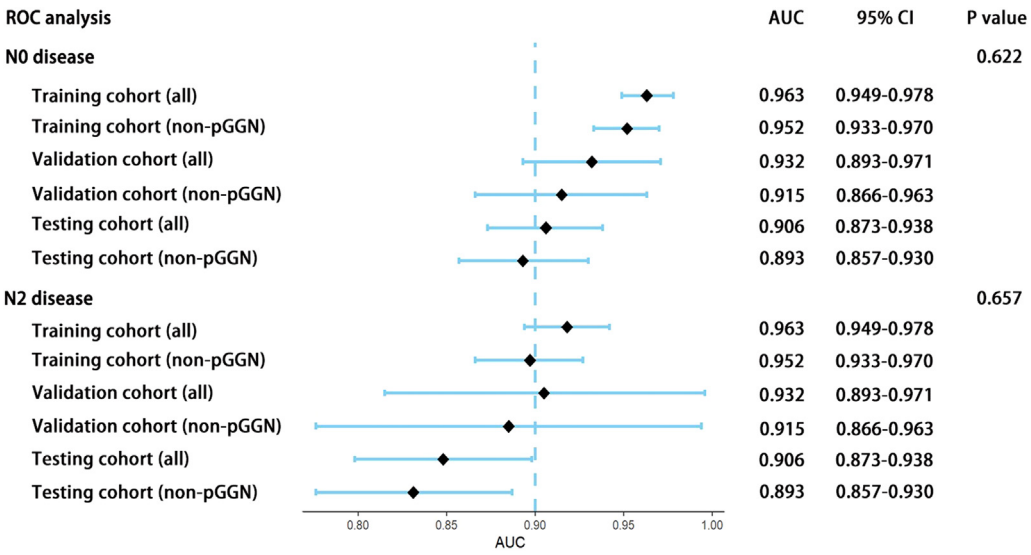

Fig. 3. ROC analysis to evaluate the model for N0 disease and N2 disease in all patients and non-pGGN patients. The *P*-values represent the differences in the AUC between the testing cohort composed of all patients and non-pGGN patients in the model. AUC, area under the curve; CI, confidence interval; N, node; pGGN, pure ground glass nodule; ROC, receiver operating characteristic.

cohort and the non-pGGN testing cohort, respectively). Studies found that both progression free survival (PFS) and overall survival (OS) of stereotactic ablative radiotherapy (SABR) were non-inferior to video-assisted thoracoscopic surgical lobectomy with mediastinal LN dissection (VATS-MLND) for operable stage IA NSCLC, which challenged the concept of preferred surgery for stage I NSCLC.[33,34] The results from JCOG0802 and JCOG0804 clinical trials suggested that segmentectomy should be the standard surgical procedure for peripheral, small size (≤ 2 cm), clinical stage IA NSCLC. Both SABR and segmentectomy are aimed to preserve more lung parenchymal for extensive treatment for relapse or second primary lung cancer, to improve OS.[35,36] However, accurate pathological N0 prediction, deciding the delineation of radiotherapy targets and the extension of the LN dissection, is the bottleneck of the application of SABR and segmentectomy in clinical stage IA lung adenocarcinoma patients. This deep learning model with high precision for predicting pathological N0 disease in stage IA lung adenocarcinoma patients may promote their application in early-stage lung adenocarcinoma. Meanwhile, the high precision deep learning model may also help thoracic surgeons to reduce ineffective LN dissection.

By contrast with previous radiomic-based or deep learning-based studies, we stratified patients based on nodule consistency and predicted pathological N0 disease and pathological N2 disease simultaneously. Previous studies had demonstrated that patients with a dominant ground glass opacity component on CT almost had no LN metastasis,[17,37] which was consistent with our findings. In this study, 100% (156/156) of patients with pGGN were pathological N0 disease. The patient population may overestimate the performance of the deep learning model. So, we stratified the patients based on nodule consistency, and found that the deep learning model showed a satisfactory predictive ability in non-pGGN patients. No significant differences were detected between the deep learning model in the whole patient group and non-pGGN patients with both pN0 and pN2 diseases. Since patients with pathological N2 disease among clinical stage IA lung adenocarcinoma patients have a poor prognosis and different therapeutic options of treatment, we tried to predict N0 and N2 disease simultaneously in the same model.[38,39] We achieved a high AUC of 0.848 (95% CI: 0.798–0.898, with a recall of 0.903) in the testing cohort and 0.831 (95% CI: 0.776–0.887, with recall of 0.931) in the non-pGGN testing cohort.

This deep learning model showed better performance in predicting LN status than other common tumor-based methods.[10,40] For example, Cong and colleagues[10] proposed a model that combined radiomics, maximum diameter and spiculation to predict lymph node metastasis, whereas Liu and colleagues[40] proposed a model that combined radiomics and pleural retraction to predict lymph node metastasis. The AUCs of these two methods are 0.86 and 0.758, respectively, which are lower than our study (AUC of 0.906 in predicting pN0 disease). Unlike other studies, this deep learning model combined the primary tumor features, lung background features, and clinical features. We also included the peritumoral microenvironment by expanding ROI of the primary tumor to the surrounding lung parenchyma by three times. A previous study proved that the peritumoral volume features could improve the prediction performance.[17] The lung background was evaluated by a thoracic radiologist and dichotomous data were inputted to keep the deep learning focusing on the tumor and peritumoral area. Considering the dependence of the deep learning model on the amount of data, we replaced the 3D CT volume with 2D CT slices and processed each CT slice independently to fully fit the deep learning model. This may be the reason why the model could achieve a superior performance for predicting pathological N0 and N2 diseases. Studies also had confirmed that clinic-radiological features could improve the ability of radiomic or artificial intelligence models to predict LN status in lung cancer patients,[41] which was consistent with our study.

This study has some limitations. Firstly, since systematic LN dissection is still preferred in clinical practice, we did not conduct prospective validation. Secondly, this study did not include the features of LNs. A combination of LN-based data and the current deep learning model might achieve better performance and requires further study. Thirdly, while ResNet-34 was chosen for its enhancements in this study, it is acknowledged that other models also harbor the potential for excellent performance. Future research should consider employing a broader range of models to explore this potential further. Fourthly, although our model shows excellent performance of diagnosis on pathological N0 disease and N2 disease in this study, the wide application of this model in clinical practice requires further studies with larger samples and more centers to determine the diagnostic cut-off value. Finally, while we have determined the individual weights of clinic-radiological features in final diagnoses, interpreting each deep learning feature is challenging due

to the inherent opacity of these models. This highlights the necessity for continued research into making deep learning models more interpretable, which is crucial for improving the clarity and understanding of our developed models.

5. Conclusions

In conclusion, the deep learning model provides a non-invasive method to predict LN status in clinical T1N0M0 lung adenocarcinoma patients. The superior performance of the model based on the primary tumor will help to target the extension of LN dissection and reduce ineffective LN dissection in early-stage lung adenocarcinoma patients.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics statement

The study was conducted in compliance with the principles of the Declaration of Helsinki and approved by the Institutional Review Board Committee of Cancer Hospital of Chinese Academy of Medical Sciences (approval number: NCCN2022C-693) and informed consent was waived.

Acknowledgements

This work was supported by the National Key R&D Program of China (grant numbers: 2020AAA0109504, 2023YFC2415200), CAMS Innovation Fund for Medical Sciences (grant number: 2021-I2M-C&T-B-061), Beijing Hope Run Special Fund of Cancer Foundation of China (grant number: LC2022A22), the National Natural Science Foundation of China (grant numbers: 81971619, 81971580, 92259302, 82372053, 91959205, 82361168664, 82022036, 81971776), Beijing Natural Science Foundation (grant number: Z20J00105), Key-Area Research and Development Program of Guangdong Province (grant number: 2021B0101420005), Strategic Priority Research Program of Chinese Academy of Sciences (grant number: XDB38040200), and the Youth Innovation Promotion Association CAS (grant number: Y2021049).

Author contributions

L.Z., H.L., S.Z. collected and analyzed the data and drafted the original manuscript. X.T., M.L., S.Y., L.Z., M.W.L. and X.Z. revised and edited the manuscript. D.D., J.T. and N.W. conceptualized the research and revised the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jncc.2024.01.005.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7–33. doi:10.3322/caac.21708.
2. Li N, Tan F, Chen W, et al. One-off low-dose CT for lung cancer screening in China: a multicentre, population-based, prospective cohort study. *Lancet Respir Med*. 2022;10(4):378–391. doi:10.1016/S2213-2600(21)00560-9.
3. de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume ct screening in a randomized trial. *N Engl J Med*. 2020;382(6):503–513. doi:10.1056/NEJMoa1911793.
4. Vaghjiani RG, Takahashi Y, Eguchi T, et al. Tumor spread through air spaces is a predictor of occult lymph node metastasis in clinical stage IA lung adenocarcinoma. *J Thorac Oncol*. 2020;15(5):792–802. doi:10.1016/j.jtho.2020.01.008.
5. Rami-Porta R, Bolejack V, Crowley J, et al. The IASLC lung cancer staging project: proposals for the revisions of the T Descriptors in the forthcoming eighth edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2015;10(7):990–1003. doi:10.1097/JTO.0000000000000559.
6. Ray MA, Smeltzer MP, Faris NR, Osarogiagbon RU. Survival after mediastinal node dissection, systematic sampling, or neither for early stage NSCLC. *J Thorac Oncol*. 2020;15(10):1670–1681. doi:10.1016/j.jtho.2020.06.009.
7. Bertolaccini L, Prisciandaro E, Bardoni C, et al. Minimally invasive anatomical segmentectomy versus lobectomy in stage IA non-small cell lung cancer: a systematic review and meta-analysis. *Cancers (Basel)*. 2022;14(24):6157. doi:10.3390/cancers14246157.
8. Kawamoto N, Tsutani Y, Kamigaichi A, et al. Tumour location predicts occult N1 nodal metastasis in clinical stage I non-small-cell lung cancer. *Eur J Cardiothorac Surg*. 2023;63(2):ezac575. doi:10.1093/ejcts/ezac575.
9. Ouyang ML, Tang K, Xu MM, Lin J, Li TC, Zheng XW. prediction of occult lymph node metastasis using tumor-to-blood standardized uptake ratio and metabolic parameters in clinical N0 lung adenocarcinoma. *Clin Nucl Med*. 2018;43(10):715–720. doi:10.1097/RLU.0000000000002229.
10. Cong M, Feng H, Ren JL, et al. Development of a predictive radiomics model for lymph node metastases in pre-surgical CT-based stage IA non-small cell lung cancer. *Lung Cancer*. Jan 2020;139:73–79. doi:10.1016/j.lungcan.2019.11.003.
11. Xue Y, Chen D, Chen Y. Reporting accuracy in prediction of lymph node metastasis of lung adenocarcinoma with radiomics. *AJR Am J Roentgenol*. 2020;215(5):W60. doi:10.2214/ajr.20.23441.
12. Xie Y, Zhao H, Guo Y, et al. A PET/CT nomogram incorporating SUVmax and CT radiomics for preoperative nodal staging in non-small cell lung cancer. *Eur Radiol*. 2021;31(8):6030–6038. doi:10.1007/s00330-020-07624-9.
13. She Y, He B, Wang F, et al. Deep learning for predicting major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer: a multicentre study. *EBioMedicine*. 2022;86:104364. doi:10.1016/j.ebiom.2022.104364.
14. Zhang L, Lv L, Li L, et al. Radiomics signature to predict prognosis in early-stage lung adenocarcinoma (≤3cm) patients with no lymph node metastasis. *Diagnostics)*. 2022;12(8):1907. doi:10.3390/diagnostics12081907.
15. Liu T, Dong D, Zhao X, et al. Radiomic signatures reveal multiscale intratumor heterogeneity associated with tissue tolerance and survival in re-irradiated nasopharyngeal carcinoma: a multicenter study. *BMC Med*. 2023;21(1):464. doi:10.1186/s12916-023-03164-3.
16. He L, Huang Y, Yan L, Zheng J, Liang C, Liu Z. Radiomics-based predictive risk score: a scoring system for preoperatively predicting risk of lymph node metastasis in patients with resectable non-small cell lung cancer. *Chin J Cancer Res*. 2019;31(4):641–652. doi:10.21147/j.issn.1000-9604.2019.04.08.
17. Wang X, Zhao X, Li Q, et al. Can peritumoral radiomics increase the efficiency of the prediction for lymph node metastasis in clinical stage T1 lung adenocarcinoma on CT? *Eur Radiol*. 2019. doi:10.1007/s00330-019-06084-0.
18. Huang EP, O'Connor JPB, McShane LM, et al. Criteria for the translation of radiomics into clinically useful tests. *Nat Rev Clin Oncol*. 2023;20(2):69–82. doi:10.1038/s41571-022-00707-0.
19. He BX, Zhong YF, Zhu YB, et al. Deep learning for predicting immunotherapeutic efficacy in advanced non-small cell lung cancer patients: a retrospective study combining progression-free survival risk and overall survival risk. *Transl Lung Cancer Res*. 2022;11(4):670–685. doi:10.21037/tlcr-22-244.
20. Tian P, He B, Mu W, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics*. 2021;11(5):2098–2107. doi:10.7150/thno.48027.
21. Dong D, Fang MJ, Tang L, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol*. 2020;31(7):912–920. doi:10.1016/j.annonc.2020.04.003.
22. Yan Y, Yao XJ, Wang SH, Zhang YD. A survey of computer-aided tumor diagnosis based on convolutional neural network. *Biology (Basel)*. 2021;10(11):1084. doi:10.3390/biology10111084.
23. Zaeemzadeh A, Rahnavard N, Shah M. Norm-preservation: why residual networks can become extremely deep? *IEEE Trans Pattern Anal Mach Intell*. 2021;43(11):3980–3990. doi:10.1109/TPAMI.2020.2990339.
24. Zhang YD, Dong Z, Wang SH, et al. Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation. *Inf Fusion*. 2020;64:149–187. doi:10.1016/j.inffus.2020.07.006.
25. Zhao X, Liang YJ, Zhang X, et al. Deep learning signatures reveal multiscale intratumor heterogeneity associated with biological functions and survival in recurrent nasopharyngeal carcinoma. *Eur J Nucl Med Mol Imaging*. 2022;49(8):2972–2982. doi:10.1007/s00259-022-05793-x.
26. Huang W, Tan K, Zhang Z, Hu J, Dong S. A review of fusion methods for omics and imaging data. *IEEE/ACM Ttrans Comput Biol Bioinform*. 2023;20(1):74–93. doi:10.1109/TCBB.2022.3143900.
27. 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(2):244–285. doi:10.1097/JTO.0b013e318206a221.
28. Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol*. 2022;17(3):362–387. doi:10.1016/j.jtho.2021.11.003.
29. Goldstraw P, Chansky K, Crowley J, et al. The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2016;11(1):39–51. doi:10.1016/j.jtho.2015.09.009.
30. De Leyn P, Vansteenkiste J, Cuypers P, et al. Role of cervical mediastinoscopy in staging of non-small cell lung cancer without enlarged mediastinal lymph nodes on CT scan. *Eur J Cardiothorac Surg*. 1997;12(5):706–712. doi:10.1016/s1010-7940(97)00253-4.

31. Lu P, Sun Y, Sun Y, Yu L. The role of (18)F-FDG PET/CT for evaluation of metastatic mediastinal lymph nodes in patients with lung squamous-cell carcinoma or adenocarcinoma. *Lung Cancer*. 2014;85(1):53–58. doi:10.1016/j.lungcan.2014.04.004.
32. Pieterman RM, van Putten JW, Meuzelaar JJ, et al. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med*. 2000;343(4):254–261. doi:10.1056/NEJM200007273430404.
33. Chang JY, Mehran RJ, Feng L, et al. Stereotactic ablative radiotherapy for operable stage I non-small-cell lung cancer (revised STARS): long-term results of a single-arm, prospective trial with prespecified comparison to surgery. *Lancet Oncol*. 2021;22(10):1448–1457. doi:10.1016/S1470-2045(21)00401-0.
34. Henschke CI, Yip R, Sun Q, et al. Prospective cohort study to compare long-term lung cancer-specific and all-cause survival of clinical early stage (T1a-b; ≤20mm) NSCLC treated by stereotactic body radiation therapy and surgery. *J Thorac Oncol*. 2023 S1556-0864(23)02265-7. doi:10.1016/j.jtho.2023.10.002.
35. Saji H, Okada M, Tsuboi M, et al. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. *Lancet*. 2022;399(10335):1607–1617. doi:10.1016/S0140-6736(21)02333-3.
36. Suzuki K, Saji H, Aokage K, et al. Comparison of pulmonary segmentectomy and lobectomy: safety results of a randomized trial. *J Thorac Cardiovasc Surg*. 2019;158(3):895–907. doi:10.1016/j.jtcvs.2019.03.090.
37. Xu S, He Z, Li X, et al. Lymph node metastases in surgically resected solitary ground-glass opacities: a two-center retrospective cohort study and pooled literature analysis. *Ann Surg Oncol*. 2023;20(6):3760–3768. doi:10.1245/s10434-023-13235-7.
38. Andrews WG, Louie BE, Castiglioni M, et al. Persistent N2 after induction is not a contraindication to surgery for lung cancer. *Ann Thorac Surg*. 2022;114(2):394–400. doi:10.1016/j.athoracsur.2021.11.010.
39. Wu LL, Liang SH, Jiang F, et al. The postoperative prognosis of skip-N2 metastasis is favorable in small-cell lung carcinoma patients with pathological N2 classification: a propensity-score-adjusted retrospective multicenter study. *Ther Adv Med Oncol*. 2023;15:17588359221146134. doi:10.1177/17588359221146134.
40. Liu Y, Kim J, Balagurunathan Y, et al. Prediction of pathological nodal involvement by CT-based Radiomic features of the primary tumor in patients with clinically node-negative peripheral lung adenocarcinomas. *Med Phys*. 2018;45(6):2518–2526. doi:10.1002/mp.12901.
41. Zhao X, Wang X, Xia W, et al. A cross-modal 3D deep learning for accurate lymph node metastasis prediction in clinical stage T1 lung adenocarcinoma. *Lung Cancer*. 2020;145:10–17. doi:10.1016/j.lungcan.2020.04.014.