scientific reports

Check for updates

OPEN

Habitat-based radiomics from contrast-enhanced CT and clinical data to predict lymph node metastasis in clinical N0 peripheral lung adenocarcinoma ≤ 3 cm

Xiaoxin Huang[1], Xiaoxiao Huang[3], Kui Wang[2], Haosheng Bai[2], Bin Ye[1] & Guanqiao Jin[2]✉

This study aims to develop an integrated model combining habitat-based radiomics and clinical data to predict lymph node metastasis in patients with clinical N0 peripheral lung adenocarcinomas measuring ≤ 3 cm in diameter. We retrospectively analyzed 1132 patients with lung adenocarcinoma from two centers who underwent surgical resection with lymph node dissection and had preoperative computed tomography (CT) scans showing peripheral nodules ≤ 3 cm. Multivariable logistic regression was employed to identify independent risk factors for the clinical model. Radiomics and habitat models were constructed by extracting and analyzing radiomic features and habitat regions from contrast-enhanced CT images. Subsequently, a combined model was developed by integrating habitat-based radiomic features with clinical characteristics. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC). The habitat model exhibited promising predictive performance for lymph node metastasis, outperforming other standalone models with AUCs of 0.962, 0.865, and 0.853 in the training, validation, and external test cohorts, respectively. The combined model demonstrated superior discriminative ability, achieving the highest AUCs of 0.983, 0.950, and 0.877 for the training, validation, and external test cohorts, respectively. The integration of habitat-based radiomic features with clinical data offers a non-invasive approach to assess the risk of lymph node metastasis, potentially supporting clinicians in optimizing patient management decisions.

Keywords Habitat imaging, Radiomics, Peripheral lung adenocarcinomas, Lymph node metastasis

Lung cancer remains the most prevalent malignancy globally, constituting approximately 12.4% of all cancer cases and 18.7% of cancer-related deaths[1]. Adenocarcinoma, the predominant histological subtype of lung cancer, has shown a rising incidence in recent years[2,3]. Lymph node metastasis is a pivotal factor in the progression of lung adenocarcinoma, significantly influencing patient prognosis. Although computed tomography (CT) is widely available and provides detailed anatomical insights, its ability to detect occult LNM, particularly in normal-sized lymph nodes, is limited, often leading to missed diagnoses[4]. Positron emission tomography/computed tomography (PET/CT) enhances detection by integrating metabolic and anatomical imaging to identify metabolically active lymph nodes. However, its sensitivity diminishes for small lesions and false positives may arise due to inflammatory conditions[5]. Endobronchial and endoscopic ultrasound offer minimally invasive sampling of mediastinal and hilar lymph nodes with high sensitivity and specificity, particularly for N2 staging[6]. Nonetheless, their limited access to certain N1 may result in undetected metastases. Mediastinoscopy, a traditional yet invasive staging method, is increasingly being supplanted and similarly struggles to evaluate all N1 stations[7]. These diagnostic limitations in assessing N1 stations can lead to understaging, suboptimal treatment planning, and adverse patient outcomes.

Surgical resection remains the cornerstone of treatment for early-stage peripheral lung adenocarcinoma. In patients with clinically N0 early-stage non-small cell lung cancer (NSCLC), postoperative pathological examination reveals lymph node upstaging to N1 or N2 in 10-20% of cases[8–10]. Controversy persists regarding

[1]Department of Radiology, Jiangbin Hospital of Guangxi Zhuang Autonomous Region, Nanning 530021, Guangxi, China. [2]Medical Imaging Center, Guangxi Medical University Cancer Hospital, Nanning 530021, Guangxi, China. [3]Department of Radiology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise 533000, Guangxi, China. ✉email: jinguanqiao77@gxmu.edu.cn

the relative benefits of systematic versus selective lymph node dissection on survival and recurrence rates in early-stage lung cancer[11,12]. While systematic lymph node dissection may enhance staging accuracy, it increases the risk of complications and potential overtreatment[13]. Conversely, selective lymph node dissection minimizes surgical trauma but risk overlooking metastatic lymph nodes. Thus, accurate preoperative prediction of lymph node status is essential for tailoring individualized treatment strategies that balance surgical risks with therapeutic efficacy.

Radiomics, the extraction of quantitative features from medical images, enables a detailed analysis of tumor heterogeneity and biological characteristics. These interpretable features can uncover malignant tumors behaviors, facilitating non-invasive clinical decision-making. Numerous studies have validated the efficacy of radiomics models in predicting lymph node metastasis in lung adenocarcinoma[14,15]. Within this context, "habitat" refers to distinct tumor subregions or volumes exhibiting unique imaging characteristics and heterogeneity, reflective of diverse biological states in the tumor microenvironment. Segmenting these habitats provides deeper insights into the tumor's internal structure and functional status[16]. Unlike traditional radiomics, which analyzes the entire tumor, habitats-based approaches partition the tumor into subregions with discrete properties, enabling a more granular assessment of spatial heterogeneity[17]. To date, habitat-based radiomics has been employed to predict treatment response, tumor aggressiveness, and genetic mutations in lung cancer[17–19]. Yet its potential for predicting lymph node metastasis remains underexplored.

In this study, we aimed to develop a predictive model for lymph node metastasis in clinical N0 peripheral lung adenocarcinoma with a diameter ≤ 3 cm using CT-derived habitat-based radiomics. We subsequently integrated these habitat features with clinical data to create a user-friendly model, seeking to improve predictive accuracy and support clinical decision-making.

Materials and methods

Patients

This retrospective study was approved by the Ethics Committee of Guangxi Medical University Cancer Hospital (approval number: KY-2022-301) and conducted in accordance with international and national ethical guidelines for biomedical research. We retrospectively enrolled patients with peripheral pulmonary nodules who underwent preoperative contrast-enhanced CT scans prior to tissue sampling and surgical resection at two medical centers. Preoperative CT images and clinical data were collected for analysis. The inclusion criteria were as follows: (1) Patients with clinical stage N0 lung adenocarcinoma; (2) Maximum tumor diameter ≤ 3 cm on chest CT; (3) Histopathological confirmation of adenocarcinoma; (4) Underwent lobectomy, segmentectomy, or wedge resection combined with lymph node dissection. Exclusion criteria included: (1) Presence of distant metastasis; (2) Receipt of neoadjuvant therapy; (3) History of other malignancies. A total of 1132 patients were included in this study (Fig. 1). Patients from center A were randomly divided into training ($n = 761$) and validation cohorts ($n = 327$) in a 7:3 ratio, while patients from centers B comprised the external test cohort ($n = 44$).

Clinical characteristics included age, gender, smoking history, cytokeratin 19-fragments (CYFRA21-1), preoperative carcinoembryonic antigen (CEA) levels, preoperative carcinoma antigen 125 (CA125) levels, maximum tumor diameter, tumor density (solid, part solid, ground glass opacity), tumor location (right upper, right middle, right lower, left upper, left lower lobes), and the presence of lobulation, spiculation, pleural indentation, air bronchogram, vascular cluster sign, vacuole, and emphysema. Two radiologists, with 20 and 9 years of experience respectively, independently assessed CT features while blinded to clinicopathological data. Discrepancies were resolved through discussion to reach a consensus. Tumor staging was performed according to the 8th edition of the International Association for the Study of Lung Cancer (IASLC) guidelines.

Image acquisition, segmentation, and preprocessing

The ITK-SNAP software (version 4.2.0, http://www.itksnap.org) was used to delineate the region of interest (ROI). A standardized pulmonary window (window width: 1800 Hounsfield units [HU], level: − 500 HU) was applied to optimize nodule segmentation. A radiologist identified the target nodule and manually adjusted the ROI boundary layer by layer, blinded to the patient's clinicopathological data. During segmentation, the ROI was confined strictly to the nodule to avoid over-segmentation beyond its borders or the unintended inclusion of adjacent structures. Pulmonary vessels, air cavities, and air-filled bronchi were excluded to ensure precise and consistent nodule delineation. The study workflow is presented in Fig. 2.

To address variability in CT scans, image preprocessing was performed prior to segmentation and feature extraction to enhance the robustness of radiomic features and ensure their suitability for analysis. Standardization of CT imaging parameters was achieved through two key steps: (1) Normalization of HU values: CT image intensities were normalized to a range of [− 1000, 200], with a window level of − 400 HU and a window width of 1200 HU. This step ensured consistent intensity levels across images obtained from different devices and scanning conditions. (2) Correction of voxel spacing: Voxel spacing within volumes of interest (VOI) was standardized to 1 mm × 1 mm × 1 mm using fixed-resolution resampling with linear interpolation, balancing efficiency and spatial detail preservation.

Habitat generation

Local features within the VOI were meticulously characterized for each voxel using CT images. Nineteen local features were calculated for each voxel to capture regional attributes of the ROI using a $5 \times 5 \times 5$ moving window. Voxels were clustered via a *k*-means algorithm, with the number of cluster centers ranging from 3 to 10. The optimal number of clusters was determined using the Calinski–Harabasz (CH) Index, Silhouette Coefficient (SC), and Davies–Bouldin (DB) Index to ensure robust cluster selection. The CH Index, which measures the ratio of between-cluster dispersion to within-cluster dispersion, favors higher values for well-defined clusters. The SC assesses the similarity of each voxel to its assigned cluster relative to other clusters, with values approaching 1

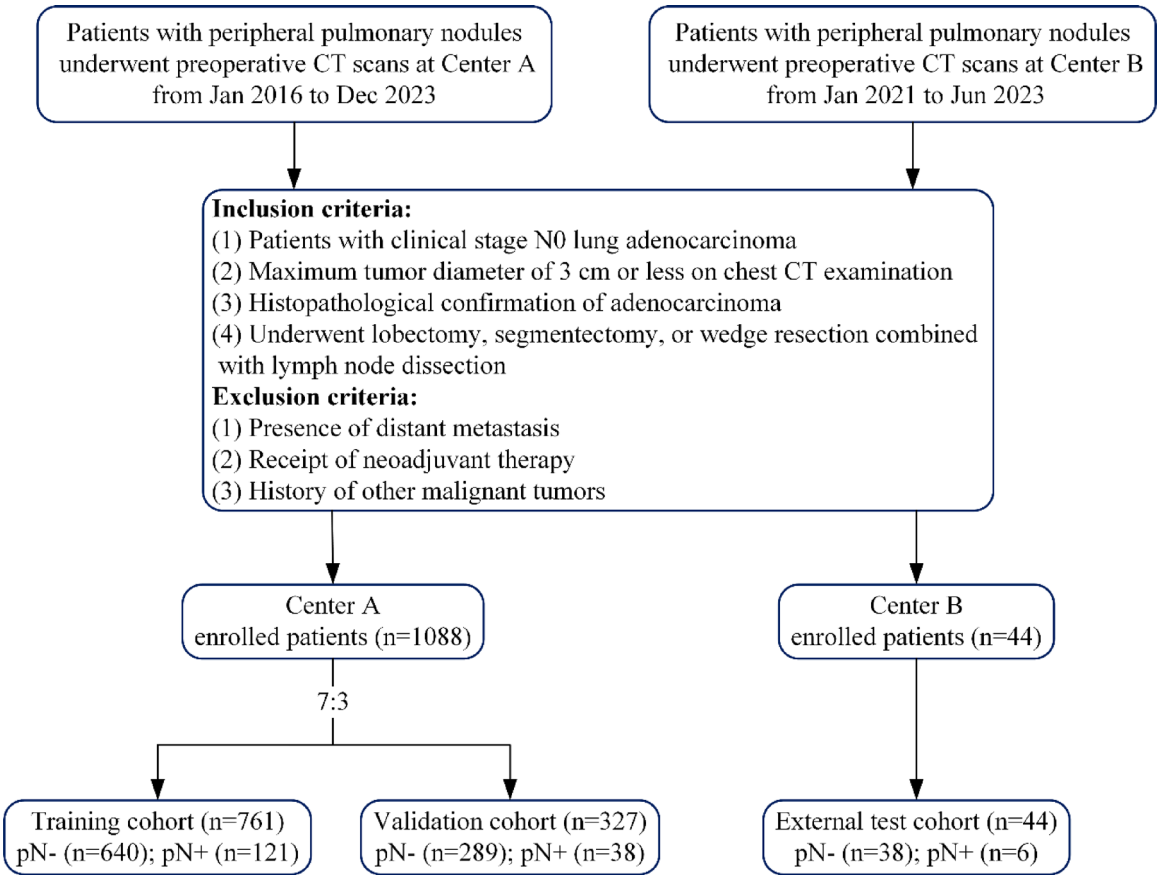

Fig. 1. Flowchart of the study subject selection. Center A, Guangxi Medical University Cancer Hospital; Center B, Affiliated Hospital of Youjiang Medical University for Nationalities.

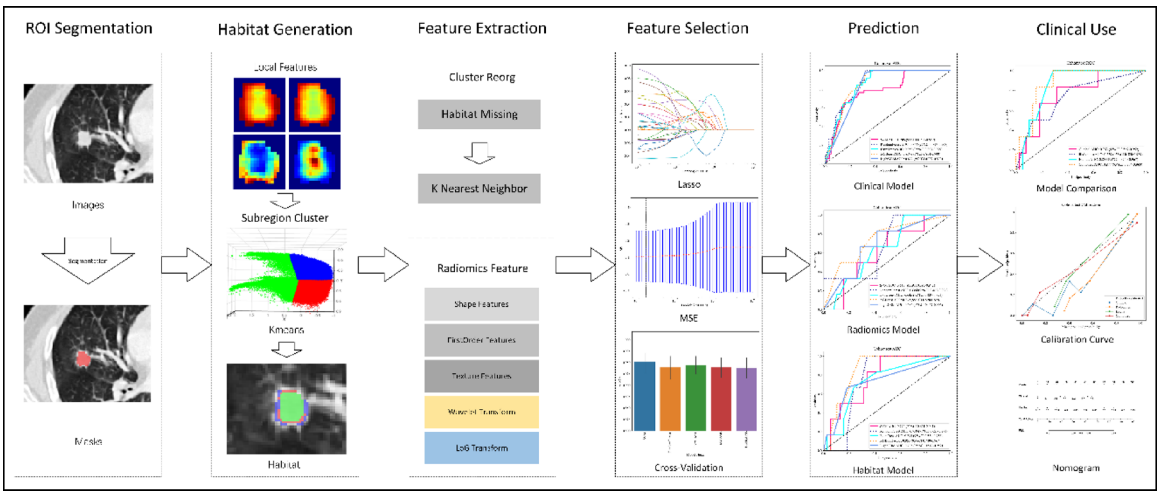

Fig. 2. Overall workflow of the work.

indicating effective clusters. The DB Index quantifies the ratio of intra-cluster to inter-cluster distances, where lower values denote superior clustering performance and greater inter-cluster separation. This methodology facilitated precise VOI segmentation and detailed evaluation of intratumor heterogeneity, providing critical insights into the tumor microenvironment for predicting lymph node metastasis. The resulting habitat regions are depicted in Fig. 3, with additional details available in Supplementary Data 1.

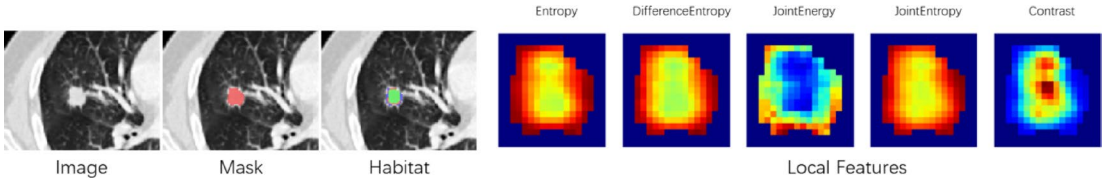

Fig. 3. Generated habitat regions.

Feature extraction

Radiomic features were systematically extracted from distinct tumor subregions and categorized into geometric, intensity, and texture categories. Geometric features delineate the tumor's shape, intensity features assess the brightness levels of voxels, and texture features, derived using techniques including Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM), capture spatial patterns inherent to each subregion.

For intratumor analysis, each subregion within the volume of VOI was identified and evaluated. In cases of incomplete clustering by the unsupervised algorithm, the K-nearest neighbors (KNN) algorithm was employed to impute missing features, ensuring consistency feature representation across all habitat regions. A detailed description of the KNN imputation process was provided in Supplementary Data 2. Following feature extraction, a feature fusion approach was implemented to integrate the distinct feature sets from each subregion prior to analysis. This process was performed using the pyradiomics tool (version 3.0.1, http://pyradiomics.readthedocs.io), adhering to the Imaging Biomarker Standardization Initiative (IBSI) guidelines to ensure standardized and reproducible feature extraction.

Feature selection

To ensure an unbiased comparison of features, Z-score normalization was applied to standardize the measurement scales across all features. P values for imaging features were computed using a t-test, with features exhibiting a P-value < 0.05 retained for further analysis. Pearson's correlation analysis was conducted to evaluate the repeatability and reliability of the features, with a focus on correlations exceeding 0.9. To address multicollinearity, highly correlated feature pairs were systematically pruned by removing one feature from each pair through a rigorous elimination process. The minimum Redundancy Maximum Relevance (mRMR) algorithm was implemented to reduce overfitting by optimizing feature selection. Within the feature selection framework, Least Absolute Shrinkage and Selection Operator (LASSO) regression played a critical role in eliminating non-essential features by shrinking their coefficients to zero. The optimal regularization parameter (λ) was determined through 10-fold cross-validation, ensuring that only the most informative and relevant features were retained for subsequent modeling.

Model and nomogram construction

The habitat and radiomics models were constructed using the final set of selected features, while the clinical model was developed based on independent predictors identified through multivariate logistic analysis. Several widely recognized machine learning algorithms were employed for model development, including support vector machine (SVM), random forest (RF), extremely randomized trees (Extra Trees), extreme gradient boosting (XGBoost), and light gradient boosting machine (Light GBM). For each model, optimal hyperparameters were identified through five-fold cross-validation in conjunction with a grid-search algorithm. A combined model was subsequently formulated, integrating insights from both the habitat and clinical models into a cohesive framework, which was visually represented through a nomogram.

Performance evaluation

The diagnostic performance of the models was evaluated using receiver operating characteristic (ROC) curves. The Delong test was applied to assess differences in predictive performance between the models. In addition, key metrics including accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to provide a comprehensive evaluation. Model calibration was examined using calibration curves and the Hosmer–Lemeshow (HL) test to assess goodness of fit. Furthermore, decision curve analysis (DCA) was performed to evaluate the clinical utility of the predictive models.

Statistical analysis

Continuous variables were presented as mean ± standard deviation (SD) and compared using the Mann–Whitney U test. Categorical variables were analyzed with either the Chi-square (χ^2) test or Fisher's exact test, as appropriate. Variables demonstrating a p-value < 0.05 in the univariate regression analysis were subsequently included in the multivariate regression analysis. All statistical analyses were conducted using IBM SPSS (version 27.0). Two-sided statistical tests were employed throughout, with statistical significance defined as $P < 0.05$.

Results

Clinical features of patients

Our study included a total of 1132 patients with clinical N0 peripheral lung adenocarcinoma, all with a diameter of 3 cm or less. Nodules sizes ranged from 3 mm to 30 mm, with a mean diameter of 16.82 ± 6.31 mm. The

cohort consisted of 967 pN- cases and 165 pN + cases. The clinical characteristics of the patients are presented in Table 1. Univariate logistic regression analysis (Table 2) identified several significant predictors of lymph node metastasis($P < 0.05$), including age, gender, maximum tumor diameter, smoking history, lung lobes, density, lobulation, spiculation, pleural indentation, air bronchogram, vascular cluster sign, vacuole, emphysema, and CYFRA21-1 levels. Multivariate regression analysis indicated that age (odds ratio [OR] = 0.954, $p < 0.001$), density (OR = 0.209, $p < 0.001$), lobulation (OR = 9.083, $p < 0.001$), pleural indentation (OR = 1.959, $p = 0.016$), and air bronchogram (OR = 0.461, $p = 0.003$) were independent predictors of lymph node metastasis.

Subregion clusters and features selection

Three subregion clustering exhibited superior performance for CT data based on CH Index, SC, and DB Index (Fig. 4). A total of 1834 unique handcrafted radiomic features were extracted and categorized into three categories: shape, first-order, and texture features. Specifically, the dataset included 14 shape features, 360 first-order features, and 1460 texture features. Figure 5a provides a graphical representation of the distribution of these feature categories, visually summarizing their allocation across the dataset.

Features with non-zero coefficients were identified using the LASSO method, with optimal lambda values determined for each model. For the habitat model, a lambda value of 0.0018 yielded the best performance, resulting in the selection of 28 features. For the radiomics model, a lambda value of 0.0083 was found to be optimal, leading to the retention of 20 features. Figure 5b–d illustrate the habitat model coefficients and mean squared error (MSE) derived from ten-fold cross-validation, alongside a histogram depicting intratumor heterogeneity based on the selected habitat features.

Performance and comparison of models

Based on the analysis of predictive performance (Supplementary Data 3), the radiomics and habitat models were constructed using XGBoost, and the clinical model was developed using Random Forest.

A summary of the predictive performance of the clinical, radiomics, habitat, and combined models is presented in Table 3; Fig. 6a–c. In the training cohort, several models demonstrated robust area under the receiver operating characteristic curve (AUC) values. The combined model achieved the highest AUC of 0.983 (95% CI 0.973–0.993), followed closely by the habitat model with an AUC of 0.962 (95% CI 0.943–0.982). The radiomics and clinical models also exhibited notable AUC values of 0.907 (95% CI 0.876–0.938) and 0.900 (95% CI 0.877–0.923), respectively. In the validation cohort, the combined model maintained strong performance with an AUC of 0.950 (95% CI 0.922–0.978). while the radiomics model also performs well, achieving an AUC of 0.908 (95% CI 0.859–0.958). The habitat and clinical models showed competitive AUC values of 0.865 (95% CI 0.798–0.932) and 0.864 (95% CI 0.819–0.908), respectively. In the external test cohort, the combined model continued to exhibit solid performance with an AUC of 0.877 (95% CI 0.764–0.990), and the habitat model followed with an AUC of 0.853 (95% CI 0.740–0.967). The clinical and radiomics models recorded AUC values of 0.763 (95% CI 0.567–0.959) and 0.739 (95% CI 0.520–0.959), respectively.

Calibration curves of the combined model displayed strong alignment between observed and predicted probabilities (Fig. 6d–f), with the HL test showing no significant deviations in the training ($P = 0.840$), validation ($P = 0.567$), and test cohort ($P = 0.860$), indicating good model calibration. The Delong test was applied to compare the AUC values across models (Fig. 6g-i). In the validation cohort, the combined exhibited a significant improvement over the clinical, radiomics, and habitat models ($P < 0.05$). DCA revealed that the combined model provided substantial net benefit based on predicted probabilities, underscoring its potential clinical utility (Fig. 7). Additionally, a nomogram was constructed to visually represent the results of the combined model (Fig. 8).

Discussion

This study introduced a comprehensive approach that integrated habitat-based radiomics with clinical data to evaluate the risk of lymph node metastasis in patients with clinical N0 peripheral lung adenocarcinoma and tumor diameter ≤ 3 cm. By combining these features, the research provided deeper insights into the association between tumor microenvironments and metastatic potential. Additionally, the study developed a user-friendly nomogram for personalized risk assessment, advancing the understanding of lymph node metastasis in early-stage lung adenocarcinoma. Moreover, the model offers a non-invasive tool for evaluating lymph node metastasis risk, with the potential to enhance clinical decision-making and improve patient management.

Radiomics, a high-throughput feature extraction technique applied to ROIs in medical images, has been widely used in studies predicting lymph node metastasis in lung adenocarcinoma based on primary tumor characteristics[20]. A previous study comparing the efficacy of radiomic features derived from intratumoral, peritumoral, and lymph nodes regions found that intratumoral features exhibited superior diagnostic performance, achieving an external validation AUC of 0.74 (95% CI 0.60–0.88)[21]. This finding underscored the significant predictive value of intratumoral radiomic features for lymph node metastasis. Radiomics leverages intratumoral heterogeneity analysis to identify features strongly correlated with metastatic potential[22]. However, the underlying biological mechanisms driving these associations remain incompletely understood. In the present study, we manually extracted 1834 radiomic features from primary lung tumors, categorized into three main groups: shape, first-order, and texture features. Of these, 20 radiomic features were significantly associated with pathologically confirmed lymph node involvement. The XGBoost machine learning algorithm achieved the best model performance, consistent with findings reported by Liu et al.[23] XGBoost's advantages over other machine learning algorithms may stem from its robust handling of missing data, prevention of overfitting, support for parallel computation, provision of feature importance scores, and ability to capture complex nonlinear relationships. These attributes suggested its potential for applications requiring accurate and robust predictive

Characteristics	Training cohort ($n=761$)		Validation cohort($n=327$)		Test cohort($n=44$)		
	pN-($n=640$)	pN+ ($n=121$)	pN-($n=289$)	pN+ ($n=38$)	pN- ($n=38$)	pN+ ($n=6$)	*P*-value
Age[a]	57.1 ± 10.4	57.8 ± 10.0	58.0 ± 9.6	54.7 ± 9.2	59.3 ± 7.8	54.2 ± 7.6	0.720
Gender							0.592
Male	231(36.1)	66(54.5)	121(41.9)	13(34.2)	27(71.1)	0(0)	
Female	409(63.9)	56(45.5)	168(58.1)	25(65.8)	11(28.9)	6(100.0)	
Maximum diameter[a]	15.9 ± 6.2	20.5 ± 5.4	16.3 ± 6.3	21.3 ± 4.4	18.8 ± 6.5	21.0 ± 6.1	0.540
Smoking history							0.597
No	524(81.9)	78(64.5)	233(80.6)	31(81.6)	31(81.6)	6(100.0)	
Yes	116(18.1)	43(35.5)	56(19.4)	7(18.4)	7(18.4)	0(0)	
Density							0.754
Solid	252(39.4)	110(90.9)	115(39.8)	37(97.4)	25(65.8)	6(100.0)	
Part solid	179(28.0)	11(9.1)	77(26.6)	1(2.6)	6(15.8)	0(0)	
Ground glass opacity	209(32.7)	0(0)	97(33.6)	0(0)	7(18.4)	0(0)	
Lung lobes							0.887
Right upper lobe	212(33.1)	39(32.2)	89(30.8)	12(31.6)	11(28.9)	2(33.3)	
Right middle lobe	63(9.8)	15(12.4)	25(8.7)	5(13.2)	2(5.3)	1(16.7)	
Right lower lobe	111(17.3)	28(23.1)	56(19.4)	9(23.7)	8(21.1)	2(33.3)	
Left upper lobe	156(24.4)	19(15.7)	72(24.9)	4(10.5)	7(18.4)	1(16.7)	
Left lower lobe	98(15.3)	20(16.5)	47(16.3)	8(21.1)	10(26.3)	0(0)	
Lobulation							0.224
Absent	362(56.6)	3(2.5)	143(49.5)	0(0)	14(36.8)	1(16.7)	
Present	278(43.4)	118(97.5)	146(50.5)	38(100.0)	24(63.2)	5(83.3)	
Spiculation							0.681
Absent	482(75.3)	35(28.9)	216(74.7)	11(28.9)	16(42.1)	3(50.0)	
Present	158(24.7)	86(71.1)	73(25.3)	27(71.1)	22(57.9)	3(50.0)	
Pleural indentation							0.107
Absent	431(67.3)	34(28.1)	170(58.8)	12(31.6)	14(36.8)	0(0)	
Present	209(32.7)	87(71.9)	119(41.2)	26(68.4)	24(63.2)	6(100.0)	
Air bronchogram							0.564
Absent	452(70.6)	82(67.8)	197(68.2)	26(68.4)	21(55.3)	3(50.0)	
Present	188(29.4)	39(32.2)	92(31.8)	12(31.6)	17(44.7)	3(50.0)	
Vascular cluster sign							0.519
Absent	598(93.4)	99(81.8)	271(93.8)	33(866.8)	32(84.2)	5(83.3)	
Present	42(6.6)	22(18.2)	18(6.2)	5(13.2)	6(15.8)	1(16.7)	
Vacuole							0.930
Absent	505(85.2)	89(73.6)	244(84.4)	27(71.1)	28(73.7)	3(50.0)	
Present	95(14.8)	32(26.4)	45(15.6)	11(28.9)	10(26.3)	3(50.0)	
Emphysema							0.848
Absent	416(65.0)	79(65.3)	188(65.1)	22(57.9)	28(73.7)	5(83.3)	
Present	224(35.0)	42(34.7)	101(34.9)	16(42.1)	10(26.3)	1(16.7)	
CEA (ng/ml)							0.668
< 5	600(93.8)	96(79.3)	276(95.5)	28(73.7)	24(89.5)	6(100.0)	
5–20	36(5.6)	17(14.0)	13(4.5)	5(13.2)	2(5.3)	0(0)	
> 20	4(0.6)	8(6.6)	0(0)	5(13.2)	2(5.3)	0(0)	
CYFRA21-1(ng/ml)							0.772
≤ 3.3	603(94.2)	112(92.6)	270(93.4)	35(92.1)	38(100.00)	6(100.0)	
> 3.3	37(5.8)	9(7.4)	19(6.6)	3(7.9)	0(0)	0(0)	
CA125 (U/mL)							0.092
< 35	619(96.7)	111(91.7)	285(98.6)	36(94.7)	38(100.00)	6(100.0)	
≥ 35	21(3.3)	10(8.3)	4(1.4)	2(5.3)	0(0)	0(0)	

Table 1. Clinical characteristics of the patients. *CEA* carcinoembryonic antigen levels, *CYFRA21-1* cytokeratin 19-fragments, *CA125* carcinoma antigen 125 levels. The p-values represented the outcomes of univariable association analysis conducted for each characteristic across three datasets. [a]Data were Mean ± SD.

Characteristics	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Age	0.972(0.969–0.974)	< 0.001*	0.954(0.938–0.969)	< 0.001*
Gender	0.347(0.311–0.386)	< 0.001*	0.604(0.386–0.945)	0.064
Maximum diameter	0.925(0.917–0.933)	< 0.001*	1.031(0.993–1.070)	0.175
Smoking history	0.371(0.276–0.497)	< 0.001*	1.039(0.613–1.761)	0.906
Lung lobes	0.597(0.563–0.634)	< 0.001*	0.938(0.823–1.068)	0.419
Density	0.054(0.034–0.088)	< 0.001*	0.209(0.123–0.355)	< 0.001*
Lobulation	0.424(0.354–0.509)	< 0.001*	9.083(4.191–19.668)	< 0.001*
Spiculation	0.544(0.436–0.678)	< 0.001*	1.676(1.082–2.596)	0.052
Pleural indentation	0.416(0.338–0.513)	< 0.001*	1.959(1.236–3.105)	0.016*
Air bronchogram	0.207(0.155–0.277)	< 0.001*	0.461(0.298–0.712)	0.003*
Vascular cluster sign	0.524(0.340–0.807)	0.014*	1.341(0.770–2.333)	0.384
Vacuole	0.337(0.241–0.471)	< 0.001*	0.975(0.614–1.548)	0.929
Emphysema	0.187(0.142–0.247)	< 0.001*	0.757(0.492–1.164)	0.287
CEA	0.803(0.576–1.119)	0.276		
CYFRA21-1	0.243(0.132–0.448)	< 0.001*	1.263(0.593–2.694)	0.611
CA125	0.476(0.253–0.896)	0.053		

Table 2. Univariate and multivariate logistic analysis of clinical characteristics. *OR* odds ratio, *95% CI* 95% confidence interval, *CEA* carcinoembryonic antigen levels, *CYFRA21-1* cytokeratin 19-fragments, *CA125* carcinoma antigen 125 levels, *Represents $p < 0.05$.

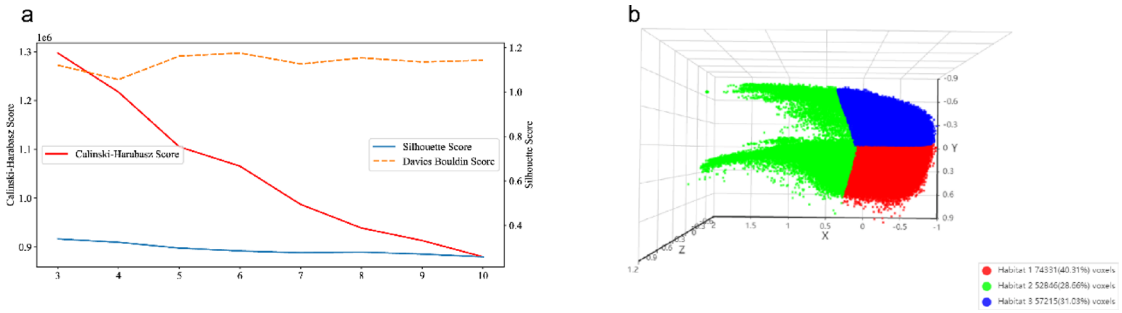

Fig. 4. (**a**) Assessment of clustering performance using the Calinski–Harabasz (CH) Index, Silhouette Coefficient (SC), and Davies–Bouldin (DB) Index across varying numbers of clusters. (**b**) Visualization of habitat features segmented into three clusters. CH Index: quantifies the ratio of between-cluster dispersion to within-cluster compactness, with higher values reflecting superior clustering quality. SC: Evaluates the compactness of each sample within its cluster relative to other clusters, with values approaching 1 indicating optimal clustering. DB Index: Evaluates the ratio of within-cluster scatter to between-cluster separation, with lower values signifying improved clustering performance.

models. In the external test cohort, the radiomics model achieved an AUC of 0.739 (95% CI 0.520–0.959), reinforcing the effectiveness of intratumoral radiomics in predicting lymph node metastasis.

Habitat imaging is a quantitative imaging technique designed to provide detailed insights into the heterogeneity of the tumor microenvironment. Previous studies have demonstrated its utility in predicting the aggressiveness of lung adenocarcinoma and distinguishing NSCLC from benign inflammatory conditions[17,24]. However, its predictive value for lymph node metastasis in clinical N0 peripheral lung adenocarcinoma remains unclear. In this study, three habitat subregions, identified via *K*-means clustering, exhibited superior performance in predicting lymph node metastasis. The number of clusters used in habitat segmentation significantly impacts feature extraction and model performance[25]. Insufficient clustering may yield overly coarse segmentation, failing to capture intratumoral heterogeneity adequately, which reduces the precision of extracted features and diminishes the model's predictive power. Conversely, excessive clusters can produce overly granular segmentation, introducing noise into the feature set and increasing model complexity, thereby elevating the risk of overfitting[26]. Through experimentation, we determined that three cluster centers struck an optimal balance between capturing tumor heterogeneity and mitigating overfitting, thereby enhancing predictive performance. This highlights the importance of selecting appropriate clustering parameters to ensure robust and reliable model outcomes. The habitat model significantly improved the prediction of lymph node metastasis in clinical N0 peripheral lung adenocarcinoma, achieving an AUC of 0.853 (95% CI 0.740–0.967), an accuracy of 0.853, and a sensitivity of 0.833 in an external test cohort. Evidence suggests that distinct tumor subregions

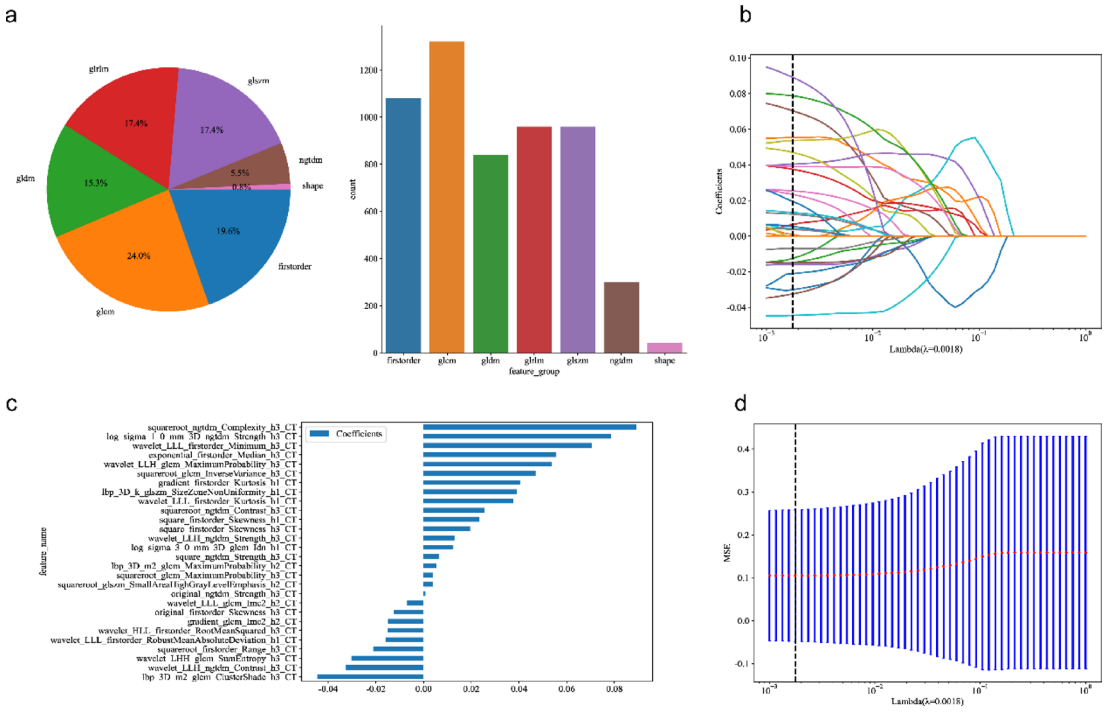

Fig. 5. (**a**) Number and proportion of handcrafted features across categories. (**b**) Coefficients derived from 10-fold cross-validation for the habitat model. (**c**) Histogram illustrating intratumor heterogeneity based on the selected habitat features. (**d**) Mean Squared Error (MSE) obtained from ten-fold cross-validation.

Models	ACC	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Cohort
Clinical	0.800	0.900	0.877–0.923	0.868	0.787	0.436	0.969	Training
Radiomics	0.873	0.907	0.876–0.938	0.810	0.884	0.570	0.961	Training
Habitat	0.917	0.962	0.943–0.982	0.868	0.927	0.691	0.974	Training
Combined	0.945	0.983	0.973–0.993	0.926	0.948	0.772	0.985	Training
Clinical	0.713	0.864	0.819–0.908	0.974	0.678	0.285	0.995	Validation
Radiomics	0.865	0.908	0.859–0.958	0.763	0.879	0.453	0.966	Validation
Habitat	0.844	0.865	0.798–0.932	0.737	0.858	0.406	0.961	Validation
Combined	0.890	0.950	0.922–0.978	0.868	0.893	0.516	0.981	Validation
Clinical	0.682	0.763	0.567–0.959	0.667	0.684	0.250	0.929	Test
Radiomics	0.682	0.739	0.520–0.959	0.667	0.684	0.250	0.929	Test
Habitat	0.773	0.853	0.740–0.967	0.833	0.763	0.357	0.967	Test
Combined	0.727	0.877	0.764–0.990	0.833	0.711	0.312	0.964	Test

Table 3. Performance comparison of different models. *ACC* accuracy, *AUC* the area under the curve, *95%CI* 95% confidence interval, *PPV* positive predictive value, *NPV* negative predictive value.

may reflect unique biological characteristics and cellular phenotypes, potentially linked to regional variations in oxygenation, vascularization, and cell density[27]. Notably, Hypoxia is a well-established driver of tumor invasion and metastasis. Under hypoxic conditions, activation of the hypoxia-inducible factor signaling pathway upregulates key genes, such as vascular endothelial growth factor (VEGF), promoting aberrant angiogenesis and enhancing tumor cell migration and invasion[28]. Furthermore, structural abnormalities in tumor vasculature lead to uneven blood perfusion, exacerbating local hypoxia and acidosis. This establishes a self-reinforcing pro-tumorigenic cycle characterized by hypoxia-induced VEGF overexpression and dysregulated angiogenesis[29]. Additionally, elevated tumor cell density may promote local invasion through compressive forces that disrupt cell-cell adhesion and activate mechanosensitive pathways, further facilitating cell migration[30]. Habitat-based radiomics enables quantitative tumor analysis by segmenting tumors into biologically meaningful subregions, thereby potentially enabling a more refined characterization of intratumoral spatial heterogeneity and revealing regional differences in perfusion, hypoxia, and metabolism[31,32]. Such subtle variations are often undetectable with conventional radiomics, which relies on features extracted from the entire tumor. Consequently, habitat imaging provides deeper insights into tumor biology and supports the noninvasive identification of aggressive

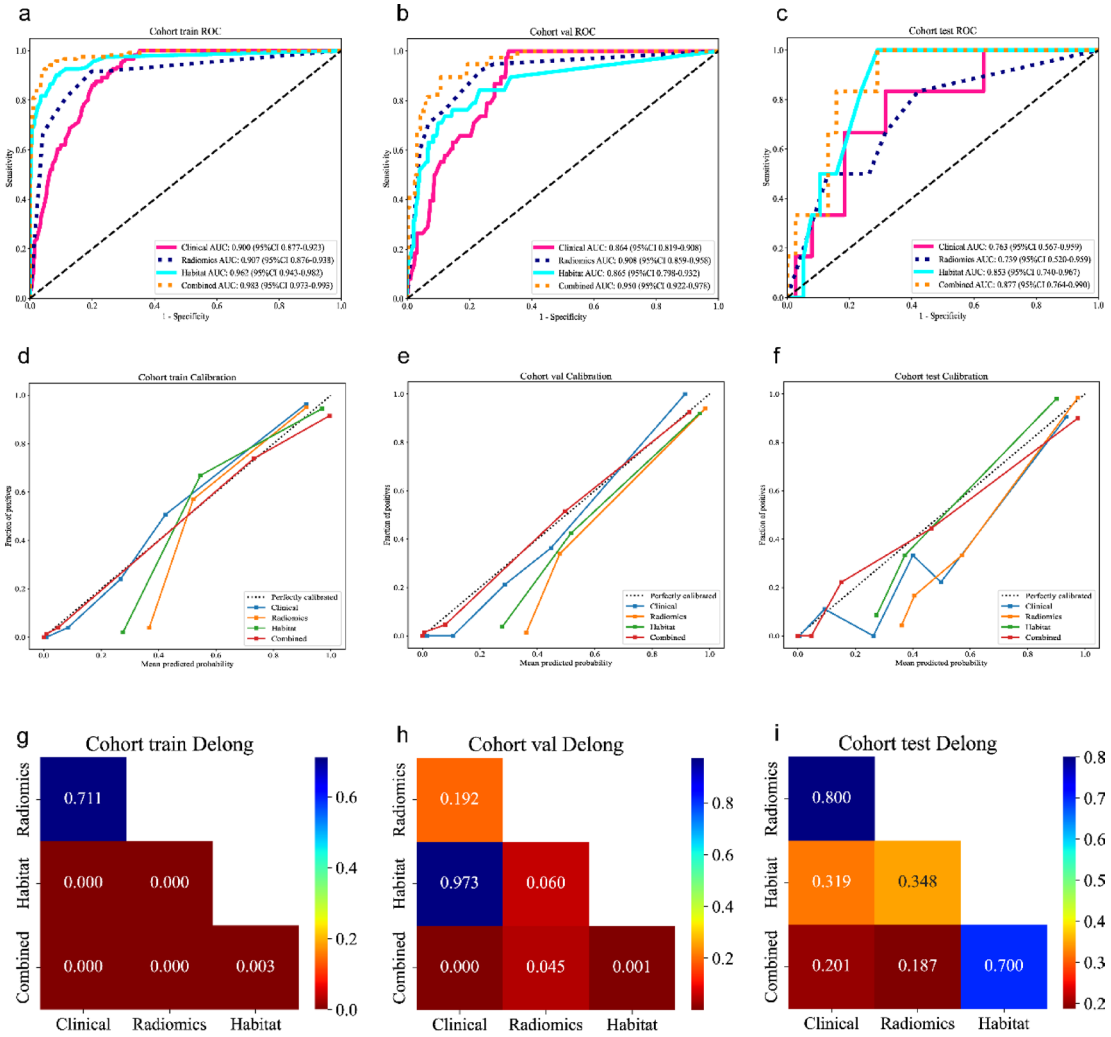

Fig. 6. Performance evaluation of different models across cohorts. (**a**–**c**) Receiver Operating Characteristic (ROC) curves for the (**a**) training, (**b**) validation, and (**c**) external test cohorts. (**d**–**f**) Calibration curves for the (**d**) training, (**e**) validation, and (**f**) external test cohorts. (**g**–**i**) Delong test results for the (**g**) training, (**h**) validation, and (**i**) external test cohorts.

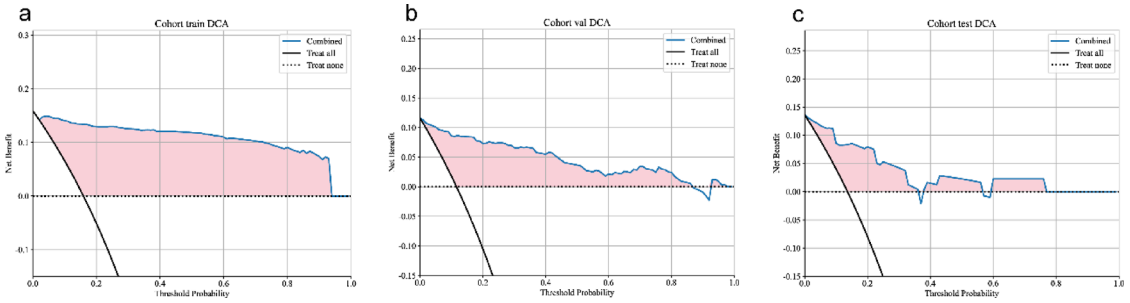

Fig. 7. Decision curve analysis (DCA) for the combined model in the (**a**) training, (**b**) validation, and (**c**) external test cohort.

and metastasis-prone subregions. This approach enhances the accuracy of metastatic risk assessment and aids in the formulation of precise, individualized treatment strategies.

Multivariate regression analyses identified several clinical variables associated with lymph node metastasis, including age, tumor density, lobulation, pleural indentation, and the presence of an air bronchogram. Many studies have explored the influence of age on lymph node metastasis in early-stage lung cancer[33–36]. Some reports

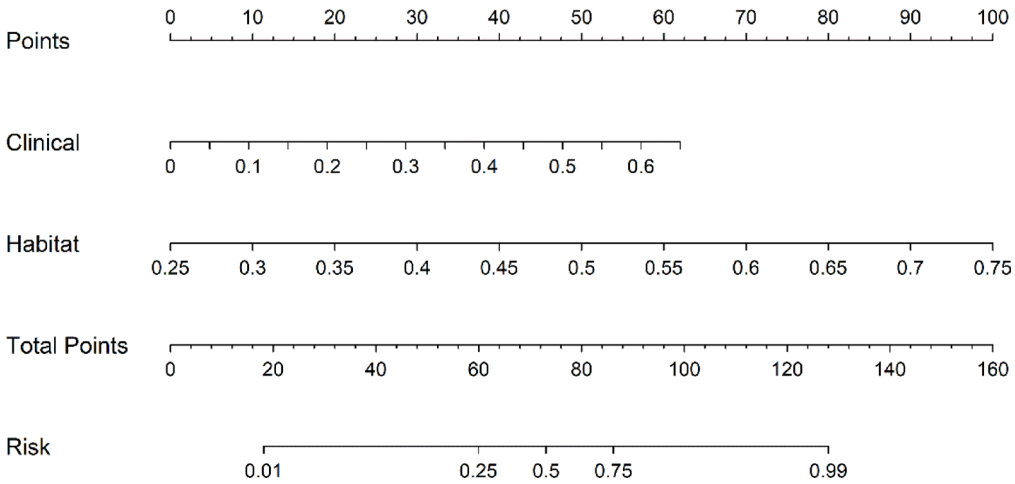

Fig. 8. Nomogram for clinical use.

suggest that younger patients face an elevated risk of lymph node metastasis[33,34], whereas others find no significant association between age and lymph node metastasis in early-stage lung adenocarcinoma[35,36]. These conflicting findings may stem from differences in sample size, patient selection criteria, and smoking status[37] across different studies. In our analysis, younger patients exhibited a higher incidence of lymph node metastasis. Additionally, lobulation and pleural indentation emerged as independent risk factors for lymph node metastasis in early-stage lung adenocarcinoma, corroborating findings by Zhao and Zhang et al.[38,39]. These imaging features reflect aggressive tumor growth and interaction with adjacent tissues, increasing the likelihood of lymph node invasion. Moreover, they may be associated with specific molecular characteristics of the tumor[40,41], further amplifying its metastatic potential. Our study also revealed a correlation between density and lymph node metastasis, with solid tumors demonstrating a heightened metastasis risk, consistent with Liu's observations[42]. Furthermore, the presence of an air bronchogram was identified as a significant indicator for predicting lymph node metastasis in adenocarcinoma. An air bronchogram, characterized by visible airway structures on CT imaging, reflects a distinct tumor growth pattern. Specifically, tumors exhibiting a lining growth pattern tend to preserve alveoli and bronchioles, indicative of surface expansion rather than invasive destruction of lung tissue[43]. Consequently, the presence of an air bronchogram may suggest more localized tumor growth, correlating with a reduced risk of lymph node metastasis. Notably, compared to standalone models, the combined model integrating habitat features with clinical data achieved higher AUC values across diverse cohorts, demonstrating superior predictive performance. This indicates that the combined approach provides more stable and reliable risk stratification in various clinical contexts, thereby enhancing its practical utility.

Overfitting remains a prevalent challenge in model development, particularly when the training cohort yields a high AUC[44]. To address this risk, we implemented feature selection, regularization, cross-validation, and external validation strategies. Despite these efforts, the notably high AUC in the training cohort (0.983) suggests a potential for overfitting, possibly due to the model capturing patterns overly specific to the training data[45]. Nevertheless, the robust AUC in the external validation cohort (0.877) demonstrates that the model retains strong generalization and is not excessively dependent on the training dataset. Comparative analysis across models further corroborates this finding. The radiomics-only model exhibited a substantial drop in AUC from internal validation (0.908) to external test (0.739), indicating a higher degree of overfitting. In contrast, the habitat model displayed greater stability, with only a minor decline from 0.865 to 0.853 in external validation, underscoring its robustness. Notably, the combined model outperformed both individual models, achieving the highest external validation AUC (0.877) while maintaining well-calibrated predictions. The HL test revealed no significant deviations ($P > 0.05$) across all cohorts, confirming reliable probability estimates and supporting its clinical utility. Although the training AUC was high (0.983), the relatively small decline in AUC across datasets indicates that the model effectively balances predictive power and overfitting risk.This study has several limitations. First, as a retrospective analysis, it is susceptible to selection bias, and complete control over all potential confounding factors was not achievable, which may compromise the reliability of the findings. Second, the study exclusively focuses on peripheral lung adenocarcinomas with a diameter of ≤ 3 cm. Despite being a dual-center study, the limited number of cases with postoperative lymph node metastasis may diminish the model's predictive accuracy and overall robustness in evaluating lymph node metastasis risk. Third, the relatively small sample size of the external test cohort ($n = 44$) may constrain the model's generalizability and robustness. A limited external test cohort can introduce statistical variability, potentially reducing the reliability of performance metrics and leading to either overestimation or underestimation of the model's true predictive capability. Additionally, the analysis was restricted to CT imaging data, without incorporating PET/CT imaging, which may limit a more holistic evaluation of lymph node metastasis. Future studies prioritize prospective validation with larger, multi-center datasets to strengthen the model's robustness and generalizability. Furthermore, the incorporation of supplementary biomarkers, such as inflammatory indices, circulating tumor DNA profiles, microRNA signatures, and protein-based markers, along with PET/CT imaging data, will be investigated to

enhance predictive accuracy and enable a more comprehensive evaluation of lymph node metastasis risk in peripheral lung adenocarcinoma.

Conclusion

In this study, a contrast-enhanced CT image-based model was developed to predict lymph node metastasis in patients with clinical N0 peripheral lung adenocarcinomas and measuring 3 cm or less in diameter. The habitat-based radiomic model outperformed other individual models in detecting subtle imaging details. A combined model, integrating the habitat model with clinical features and presented as a nomogram, was created, demonstrating its feasibility and efficiency in predicting lymph node metastasis. This contrast-enhanced CT-based nomogram offers a non-invasive, cost-effective approach with potential for broad clinical adoption, aiding physicians in making more accurate and informed clinical decisions.

Data availability

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

Received: 30 October 2024; Accepted: 12 May 2025
Published online: 16 May 2025

References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **74**, 229–263. https://doi.org/10.3322/caac.21834 (2024).
2. Leiter, A., Veluswamy, R. R. & Wisnivesky, J. P. The global burden of lung cancer: current status and future trends. *Nat. Rev. Clin. Oncol.* **20**, 624–639. https://doi.org/10.1038/s41571-023-00798-3 (2023).
3. Barta, J. A. & Powell, CAWisnivesky, J. P. Global epidemiology of lung Cancer. *Ann. Glob. Health.* **85**, 8. https://doi.org/10.5334/aogh.2419 (2019).
4. Beigelman-Aubry, C., Dunet, V. & Brun, A. L. CT imaging in pre-therapeutic assessment of lung cancer. *Diagn. Interv. Imaging.* **97**, 973–989. https://doi.org/10.1016/j.diii.2016.07.010 (2016).
5. Park, H. K. et al. Occult nodal metastasis in patients with non-small cell lung cancer at clinical stage IA by PET/CT. *Respirology* **15**, 1179–1184. https://doi.org/10.1111/j.1440-1843.2010.01793.x (2010).
6. Al-Ibraheem, A. et al. Impact of 18F-FDG PET/CT, CT and EBUS/TBNA on preoperative mediastinal nodal staging of NSCLC. *BMC Med. Imaging.* **21**, 49. https://doi.org/10.1186/s12880-021-00580-w (2021).
7. Dunne, E. G., Fick, C. N. & Jones, D. R. Mediastinal staging in non-small-cell lung cancer: saying goodbye to mediastinoscopy. *J. Clin. Oncol.* **41**, 3785–3790. https://doi.org/10.1200/jco.23.00867 (2023).
8. , G. Ghaly et al. Clinical predictors of nodal metastases in peripherally clinical T1a N0 non-small cell lung cancer. *Ann. Thorac. Surg.* **104**, 1153–1158. https://doi.org/10.1016/j.athoracsur.2017.02.074 (2017).
9. Ismail, M. et al. Lymph node upstaging for non-small cell lung cancer after uniportal video-assisted thoracoscopy. *J. Thorac. Dis.* **10**, S3648–S3654. https://doi.org/10.21037/jtd.2018.06.70 (2018).
10. Licht, P. B., Jørgensen, O. D., Ladegaard, L. J. & Akobsen, E. A. National study of nodal upstaging after thoracoscopic versus open lobectomy for clinical stage i lung cancer. *Ann. Thorac. Surg.* **96**, 943–950. https://doi.org/10.1016/j.athoracsur.2013.04.011 (2013).
11. Cheng, X. et al. Impact of lymph node dissection on survival and tumor recurrence for patients with resected cT1-2N0 small cell lung Cancer. *Ann. Surg. Oncol.* **29**, 7512–7525. https://doi.org/10.1245/s10434-022-12215-7 (2022).
12. Adachi, H. et al. Lobe-specific lymph node dissection as a standard procedure in surgery for non-small cell lung cancer: A propensity score matching study. *J. Thorac. Oncol.* **12**, 85–93. https://doi.org/10.1016/j.jtho.2016.08.127 (2017).
13. Luo, J., Yang, S. & Dong, S. Selective mediastinal lymphadenectomy or complete mediastinal lymphadenectomy for clinical stage I Non-Small cell lung cancer: A meta-analysis. *Adv. Ther.* **38**, 5671–5683. https://doi.org/10.1007/s12325-021-01954-w (2021).
14. Xue, Y., Chen, D. & Chen, Y. Reporting accuracy in prediction of lymph node metastasis of lung adenocarcinoma with radiomics. *Am. J. Roentgenol.* **215**, W60. https://doi.org/10.2214/ajr.20.23441 (2020).
15. Liu, 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.* **45**, 2518–2526. https://doi.org/10.1002/mp.12901 (2018).
16. Sala, E. et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin. Radiol.* **72**, 3–10. https://doi.org/10.1016/j.crad.2016.09.013 (2017).
17. Shang, Y. et al. Habitat imaging with tumoral and peritumoral radiomics for prediction of lung adenocarcinoma invasiveness on preoperative chest CT: A multicenter study. *Am. J. Roentgenol.* **223**, e2431675. https://doi.org/10.2214/ajr.24.31675 (2024).
18. Wu, J. et al. Habitat radiomics and deep learning fusion nomogram to predict EGFR mutation status in stage I non-small cell lung cancer: a multicenter study. *Sci. Rep.* **14**, 15877. https://doi.org/10.1038/s41598-024-66751-1 (2024).
19. Cherezov, D. et al. Revealing tumor habitats from texture heterogeneity analysis for classification of lung Cancer malignancy and aggressiveness. *Sci. Rep.* **9**, 4500. https://doi.org/10.1038/s41598-019-38831-0 (2019).
20. Wu, T. et al. Predictive value of radiomic features extracted from primary lung adenocarcinoma in forecasting thoracic lymph node metastasis: a systematic review and meta-analysis. *BMC Pulm. Med.* **24**, 246. https://doi.org/10.1186/s12890-024-03020-x (2024).
21. Das, S. K. et al. Integrative nomogram of intratumoral, peritumoral, and lymph node radiomic features for prediction of lymph node metastasis in cT1N0M0 lung adenocarcinomas. *Sci. Rep.* **11**, 10829. https://doi.org/10.1038/s41598-021-90367-4 (2021).
22. Wen, X. et al. CT-based radiomic consensus clustering association with tumor biological behavior in clinical stage IA adenocarcinoma: a retrospective study. *Translational Lung Cancer Res.* **13**, 1794–1806. https://doi.org/10.21037/tlcr-24-283 (2024).
23. Liu, M. W. et al. A comparison of machine learning methods for radiomics modeling in prediction of occult lymph node metastasis in clinical stage IA lung adenocarcinoma patients. *J. Thorac. Dis.* **16**, 1765–1776. https://doi.org/10.21037/jtd-23-1578 (2024).
24. Chen, L. et al. Habitat imaging-based 18F-FDG PET/CT radiomics for the preoperative discrimination of non-small cell lung cancer and benign inflammatory diseases. *Front. Oncol.* **11**, 759897. https://doi.org/10.3389/fonc.2021.759897 (2021).
25. Zhu, Y. et al. Prediction of therapeutic response to transarterial chemoembolization plus systemic therapy regimen in hepatocellular carcinoma using pretreatment contrast-enhanced MRI based habitat analysis and crossformer model. *Abdom. Radiol.* https://doi.org/10.1007/s00261-024-04709-7 (2024).
26. Kachouie, N., Deebani, N., Shutaywi, W. & Christiani, M. Lung cancer clustering by identification of similarities and discrepancies of DNA copy numbers using maximal information coefficient. *PLoS One.* **19**, e0301131. https://doi.org/10.1371/journal.pone.0301131 (2024).
27. Sagreiya, H. Finding the pieces to treat the whole: using radiomics to identify tumor habitats. *Radiol. Artif. Intell.* **6**, e230547. https://doi.org/10.1148/ryai.230547 (2024).

28. Liao, C., Liu, X., Zhang, C. & Zhang, Q. Tumor hypoxia: from basic knowledge to therapeutic implications. *Sem. Cancer Biol.* **88**, 172–186. https://doi.org/10.1016/j.semcancer.2022.12.011 (2023).
29. Kaur, G. & Roy, B. Decoding tumor angiogenesis for therapeutic advancements: mechanistic insights. *Biomedicines* **12**, 827. https://doi.org/10.3390/biomedicines12040827 (2024).
30. Gao, Y., Pan, Z., Li, H., Wang, F. & Raza, F. Antitumor therapy targeting the tumor microenvironment. *J. Oncol.* https://doi.org/10.1155/2023/6886135 (2023).
31. Bailo, M. et al. Decoding the heterogeneity of malignant gliomas by PET and MRI for Spatial habitat analysis of hypoxia, perfusion, and diffusion imaging: A preliminary study. *Front. NeuroSci.* **16**, 885291. https://doi.org/10.3389/fnins.2022.885291 (2022).
32. Wu, M. et al. Predicting the early therapeutic response to hepatic artery infusion chemotherapy in patients with unresectable HCC using a contrast-enhanced computed tomography-based habitat radiomics model: a multi-center retrospective study. *Cell. Oncol.* https://doi.org/10.1007/s13402-025-01041-0 (2025).
33. Xue, M. et al. The role of adenocarcinoma subtypes and immunohistochemistry in predicting lymph node metastasis in early invasive lung adenocarcinoma. *BMC Cancer*. **24**, 139. https://doi.org/10.1186/s12885-024-11843-4 (2024).
34. Chen, B. et al. Lymph node metastasis in Chinese patients with clinical T1 non-small cell lung cancer: A multicenter real-world observational study. *Thorac. Cancer*. **10**, 533–542. https://doi.org/10.1111/1759-7714.12970 (2019).
35. Gallina, F. T. et al. ALK rearrangement is an independent predictive factor of unexpected nodal metastasis after surgery in early stage, clinical node negative lung adenocarcinoma. *Lung Cancer*. **180**, 107215. https://doi.org/10.1016/j.lungcan.2023.107215 (2023).
36. Choi, P. et al. Fascin immunoreactivity for preoperatively predicting lymph node metastases in peripheral adenocarcinoma of the lung 3 cm or less in diameter. *Eur. J. Cardiothorac. Surg.* **30**, 538–542. https://doi.org/10.1016/j.ejcts.2006.06.029 (2006).
37. Shan, L. et al. Chinese never smokers with adenocarcinoma of the lung are younger and have fewer lymph node metastases than smokers. *Respir. Res.* **23**, 293. https://doi.org/10.1186/s12931-022-02199-z (2022).
38. Zhao, F. et al. Predictability and utility of contrast-enhanced CT on occult lymph node metastasis for patients with clinical stage IA-IIA lung adenocarcinoma: A double-center study. *Acad. Radiol.* **30**, 2870–2879. https://doi.org/10.1016/j.acra.2023.03.002 (2023).
39. Zhang, W. et al. Lymph node metastasis and its risk factors in T1 lung adenocarcinoma. *Thorac. Cancer*. **14**, 2993–3000. https://doi.org/10.1111/1759-7714.15088 (2023).
40. Tian, K., Li, Z. & Qin, L. Detection of CEA and ProGRP levels in BALF of patients with peripheral lung Cancer and their relationship with CT signs. *Biomed. Res. Int.* **2022**, 4119912 https://doi.org/10.1155/2022/4119912 (2022).
41. Watari, N. et al. Characteristic computed tomography features in mesenchymal-epithelial transition exon14 skipping-positive non-small cell lung cancer. *BMC Pulm. Med.* **22**, 260. https://doi.org/10.1186/s12890-022-02037-4 (2022).
42. Liu, M. et al. Growth characteristics of early-stage (IA) lung adenocarcinoma and its value in predicting lymph node metastasis. *Cancer Imaging*. **23**, 115. https://doi.org/10.1186/s40644-023-00631-1 (2023).
43. Wang, M. et al. Clinical characterization of node-negative lung adenocarcinoma: results of a prospective investigation. *J. Thorac. Oncol.*. **1**, 825–831. https://doi.org/10.1016/S1556-0864(15)30412-3 (2006).
44. Gygi, J. P., Kleinstein, S. H. & Guan, L. Predictive overfitting in immunological applications: pitfalls and solutions. *Hum. Vaccines Immunother*. **19**, 2251830. https://doi.org/10.1080/21645515.2023.2251830 (2023).
45. Samadi, M. E., Mirzaieazar, H., Mitsos, A. & Schuppert, A. Noisecut: a python package for noise-tolerant classification of binary data using prior knowledge integration and max-cut solutions. *BMC Bioinform.* **25**, 155. https://doi.org/10.1186/s12859-024-05769-8 (2024).

Acknowledgements

We sincerely thank the Medical Imaging Center of the Guangxi Medical University Cancer Hospital, the Medical Imaging Center of the Affiliated Hospital of Youjiang Medical University for Nationalities, and the OnekeyAI platform for their invaluable assistance in this work.

Author contributions

X.X.H. and G.Q.J. conceptualized and designed this study. X.X.H., X.X.H., K.W., H.S.B., and B.Y. collected clinical and radiological data. X.X.H. and G.Q.J. conducted the data analysis and drafted the manuscript. All authors read and approved the final manuscript. X.X.H. refers to X.H., X.X.H. refers to X.H.

Funding

This study has received funding from Beijing Medical Award Foundation (YXJL-2022-0665-0210); Guangxi Key Research and Development Program (GuikeAB23026087), and Natural Science Foundation of Guangxi Zhuang Autonomous Region (2023GXNSFAA026225).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-02181-x.

Correspondence and requests for materials should be addressed to G.J.