

AI-derived longitudinal and multi-dimensional CT classifier for non-small cell lung cancer to optimize neoadjuvant chemoimmunotherapy decision: a multicentre retrospective study



Guanchao Ye,^{a,p} Zhitao Wei,^{b,c,d,p} Chu Han,^{c,d,e,p} Guangyao Wu,^{f,p} Chinting Wong,^{c,d} Yanting Liang,^{c,d} Xiaobo Chen,^{c,d} Wang Zhou,^g Jianbo Gao,^h Changhong Liang,^{c,d} Yongde Liao,ⁱ Lizza E. L. Hendriks,^j Leonard Wee,^{k,l} Dirk De Ruyscher,^{l,m} Andre Dekker,^{k,l} Haiyu Zhou,^{n,****} Yu Qi,^{a,o,***} Zaiyi Liu,^{c,d,**} and Zhenwei Shi^{c,d,e,*}



^aDepartment of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China

^bSchool of Medicine, South China University of Technology, Guangzhou, 510080, China

^cDepartment of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China

^dGuangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou, 510080, China

^eMedical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China

^fDepartment of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430000, China

^gThe First Clinical College of Zhengzhou University, Zhengzhou, 450001, China

^hDepartment of Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China

ⁱDepartment of Thoracic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430000, China

^jDepartment of Pulmonary Diseases, GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Center+, Maastricht, 6202 AZ, the Netherlands

^kClinical Data Science, Faculty of Health Medicine and Life Sciences, Maastricht University, Maastricht, 6229 ET, the Netherlands

^lDepartment of Radiation Oncology (Maastricht), GROW School of Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, 6229 ET, the Netherlands

^mDepartment of Radiotherapy, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, 3000 CA, the Netherlands

ⁿDepartment of Thoracic Surgery, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China

^oHenan Province Engineering Research Center of Translational Precision Medicine for Lung Cancer, Zhengzhou, 450000, China

Summary

Background Neoadjuvant chemoimmunotherapy (NACI) has significantly improved survival in patients with resectable non-small cell lung cancer (NSCLC). However, with the currently available methods (PD-L1, RECIST), it is difficult to predict who will benefit from treatment before therapy and who will achieve pathological complete response (pCR) before surgery. Non-invasive methods to predict treatment response to NACI could be used to further personalize treatment.

Methods In the multicenter retrospective study, we enrolled 534 patients with NSCLC who received NACI followed by surgical resection at three Chinese hospitals between January 2019 and December 2024 (193 patients from Center A, 193 patients from Center B, and 148 patients from Center C). We developed and validated Lung Cancer Neoadjuvant Immuno-Chemotherapy Response Predictor (LC-NICER), a CT-based Artificial Intelligence (AI) system that integrates longitudinal radiomics (tumor texture), deep learning (microenvironmental context), and habitat imaging (tumor and peritumoral subregional dynamics) to predict pCR by analyzing tumor spatiotemporal heterogeneity. The LC-NICER system consists of two complementary predictive models: LC-NICER_α, a pretreatment model that identifies patients likely to benefit from NACI to guide personalized therapy, and LC-

eClinicalMedicine
2025;89: 103551

Published Online xxx
<https://doi.org/10.1016/j.eclinm.2025.103551>

*Corresponding author. Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China.

**Corresponding author. Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China.

***Corresponding author. Department of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China.

****Corresponding author.

E-mail addresses: shizhenwei@gdph.org.cn (Z. Shi), liuzaiyi@gdph.org.cn (Z. Liu), qiyu@zzu.edu.cn (Y. Qi), zhouhaiyu@gdph.org.cn (H. Zhou).

^PThese authors contributed equally to this work.

NICER_s, a preoperative model that evaluates tumor regression and resection feasibility to inform surgical planning. This study is registered at [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT06285058) (NCT06285058).

Findings Among the 386 patients from Center A and B, 308 were randomly assigned to the training dataset and 78 to the internal validation dataset, following an 8:2 split ratio. The 148 patients from Center C formed an independent and external test dataset. LC-NICER prediction system demonstrated excellent performance with the area under the curves (AUCs) of 0.950 (0.927–0.970) in the training cohort, 0.889 (0.796–0.961) in the internal validation cohort, and 0.870 (0.803–0.927) in the external test cohort. The LC-NICER_α achieved an accuracy of 0.722 (0.668–0.772) before therapy, while LC-NICER_s showed significantly improved accuracy of 0.831 (0.800–0.861) before surgery. Notably, LC-NICER outperformed current clinical standards, with absolute accuracy improvements of 10% over PD-L1 testing (0.622 [0.564–0.683], $p = 0.002$) and 18% over RECIST 1.1 criteria (0.651 [0.610–0.693], $p = 0.008$). For easy clinical utility and research reproducibility, we developed and openly published a software.

Interpretation As a non-invasive AI system for predicting NACI response in NSCLC, LC-NICER may offer future clinical personalized therapeutic strategies, accelerate adaptive clinical trials, and optimize treatment decisions, potentially reducing reliance on invasive procedures.

Funding This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0531100, 2024ZD0531101), the National Natural Science Foundation of China (82472062, 62102103, 82473298, 82202148, 82502479), the National Key Research and Development Program of China (2023YFC2508603), the Natural Science Foundation of Guangdong Province of China (2024A1515011672), Regional Innovation and Development Joint Fund of National Natural Science Foundation of China (U22A20345), Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (2022B1212010011).

Copyright © 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

Keywords: Non-small cell lung cancer; Neoadjuvant chemoimmunotherapy; Radiomics; Deep learning; Habitat imaging; Longitudinal CT

Introduction

Neoadjuvant chemoimmunotherapy (NACI) or peri-operative immunotherapy has transformed non-small cell lung cancer (NSCLC) management by enabling tumor downstaging and improving disease-free survival (DFS) and overall survival (OS).^{1–5} Importantly, the depth of pathological response, and especially the achievement of pathological complete response (pCR) after NACI is correlated with improved OS.⁶ Despite the clinical benefits of NACI, approximately 40% of patients experience no therapeutic response while still being exposed to treatment-related toxicities and unnecessary healthcare costs.⁷ This highlights the critical need for accurate pre-treatment prediction of therapeutic response to optimize patient selection for NACI. Furthermore, given the prognostic significance of pCR and its association with long-term survival,⁸ a non-surgery strategy may be viable for pCR-achieving patients.^{9–11} Reliable non-invasive therapy response prediction could thus enable personalized treatment strategies, including surgery-sparing approaches for exceptional responders.

While PD-L1 expression, endorsed by the Food and Drug Administration (FDA), is the primary biomarker for immunotherapy selection,¹² its clinical utility is hampered by sampling bias, procedural complexity and limited accuracy.^{13–15} Additionally, although a higher PD-L1 expression level is correlated with an increased

chance of tumor response to immunotherapy, not all tumors with high PD-L1 respond. Notably, the NADIM trial uncovered a striking imaging-pathology discordance: 33% of tumors meeting imaging criteria for stable disease (SD) and 73% classified as partial response (PR) were pathologically confirmed as pCR following resection.¹⁶ Both CT-based (Response Evaluation Criteria in Solid Tumors [RECIST]) criteria for solid tumor correlate poorly with pathological outcomes.^{4,17} Current clinical workflow relies on invasive biopsies (PD-L1) to guide NACI decisions and traditional unidimensional RECIST criteria to evaluate clinical response. These approaches are often inefficient and inaccurate, as they fail to capture the dynamic evolution of tumors during treatment.¹⁸

CT imaging offers an objective and cost-effective approach for assessing tumor morphological and metabolic dynamics.¹⁹ Recent advances in quantitative CT imaging enable comprehensive characterization of tumor heterogeneity and microenvironmental features, providing a non-invasive alternative to tissue biopsy.^{20–22} For instance, several studies^{23,24} have demonstrated the potential of using artificial intelligence (AI) integrating pre-treatment CT data to predict major pathological response (MPR) and pCR. However, these AI approaches are constrained by their reliance on single-timepoint imaging data and small sample sizes,

Research in context

Evidence before this study

We searched PubMed from database inception to July 8, 2025, for publications on artificial intelligence-based methods to predict response to neoadjuvant chemoimmunotherapy for lung cancer. We used the search terms “artificial intelligence” or “machine learning” or “deep learning” or “radiomics”, “neoadjuvant chemoimmunotherapy” or “neoadjuvant immunochemotherapy”, “predict response”, and “lung cancer”, without language restrictions. We found 44 original studies that applied radiomics and deep learning analysis to predict pathological complete response or related attributes, mostly based on CT and PET-CT, highlighting the potential association between radiomics and deep learning features and neoadjuvant chemoimmunotherapy sensitivity. Despite encouraging preliminary results, the clinical applicability of these methods remains unclear because of small sample size. Moreover, most prior efforts focus on isolated feature types, such as radiomics or deep learning, neglecting the potential synergies of hybrid models. In addition, these AI approaches are constrained by their reliance on single-timepoint imaging data of these studies, which cannot fully reflect the temporal heterogeneity of tumors. Furthermore, no studies have bridged the gap between pre-treatment prediction and post-treatment response assessment through a unified framework. This highlights the need for a comprehensive approach that leverages longitudinal imaging data and multi-dimensional feature integration to accurately predict treatment response and guide clinical decision-making.

Added value of this study

While neoadjuvant chemoimmunotherapy (NACI) has revolutionized lung cancer management, current clinical biomarkers are unable to resolve spatially heterogeneous and temporally evolving tumor-immune interactions, with

suboptimal accuracy such as PD-L1 (0.622 [95% CI: 0.564–0.683]) and RECIST (0.651 [95% CI: 0.610–0.693]). These limitations highlight the clinical imperative for more reliable predictive tools that can non-invasively guide treatment decisions. In this study, we developed Lung Cancer Neo-adjuvant Immuno-ChEmotherapy Response predictor (LC-NICER), a spatiotemporal fusion system that combines radiomics, deep learning, and dynamic habitat imaging features from longitudinal CT data to predict treatment response, overcoming the limitations of static or single-modality approaches. The performance of LC-NICER was comprehensively evaluated through quantitative analysis, demonstrating superior predictive accuracy compared to current standard methods. As a non-invasive Artificial Intelligence (AI) tool, LC-NICER can enable personalized therapeutic strategies, accelerate adaptive clinical trials, and optimize treatment decisions, potentially reducing reliance on invasive procedures.

Implications of all the available evidence

This study introduces LC-NICER, a CT-based AI system that integrates spatiotemporal tumor heterogeneity mapping with multi-dimensional quantitative imaging features to predict treatment response to NACI. The system can stratify patients into pCR and non-pCR groups at different stages (pre- and post-treatment), serving two critical clinical purposes: (1) identification of patients who are likely to prior to NACI treatment, and (2) supporting clinical decision-making for surgery-sparing strategies following NACI treatment. Extensive validation demonstrates LC-NICER’s superior predictive accuracy compared to currently available methods (PD-L1 and RECIST), with habitat imaging analysis providing spatial mapping of metabolically active subregions associated with treatment response.

which cannot fully reflect the temporal heterogeneity of tumors. Additionally, emerging evidence emphasizes that integrating multi-timepoint imaging is critical to capturing treatment-induced biological shifts.^{25–28} Yet, most prior efforts focus on isolated feature types, such as radiomics²⁹ or deep learning,³⁰ neglecting the potential synergies of hybrid models. To our knowledge, no studies have bridged the gap between pre-treatment prediction and post-treatment response assessment through a unified framework. This highlights the need for a comprehensive approach that leverages longitudinal imaging data and multi-dimensional feature integration to accurately predict treatment response and guide clinical decision-making.

To address these gaps, we developed the Lung Cancer Neo-adjuvant Immuno-ChEmotherapy Response predictor (LC-NICER) through a multicenter study of 534 patients with longitudinal CT data. LC-NICER utilizes longitudinal CT imaging to dynamically classify patients

as treatment responders or non-responders, facilitating personalized therapy optimization and surgical planning. The system can stratify patients into pCR and non-pCR groups at different stages (before therapy and before surgery), serving two critical clinical purposes: (1) identification of patients who are likely to prior to NACI treatment, and (2) supporting clinical decision-making for surgery-sparing strategies following NACI treatment. This non-invasive longitudinal and multi-dimensional CT classifier establishes a novel paradigm for precision oncology, enabling adaptive clinical trial design and individualized therapeutic management in lung cancer.

Methods

Study design and participants

This retrospective multicenter study included patients with NSCLC who underwent surgical resection after NACI between January 2019 and December 2024 from

Union Hospital, Tongji Medical College, Huazhong University (UH, Center A, Central China), the First Affiliated Hospital of Zhengzhou University (ZZ, Center B, North China), and Guangdong Provincial People's Hospital (GDPH, Center C, South China). This cross-regional (Central, North, and South China) multi-center design is a core advantage of our study. The study protocol was registered at [ClinicalTrials.gov](https://www.clinicaltrials.gov) (NCT06285058) and approved by the institutional

review boards of all participating centers (Approval No: UCT240116). Given the retrospective nature of our study, the ethics committee exempted informed consent. The study included 193 patients from Center A, 193 patients from Center B, and 148 patients from Center C. A detailed flowchart of patient inclusion is provided in [Fig. 1](#).

Among the 386 patients from Center A and B, 308 were randomly assigned to the training dataset and 78

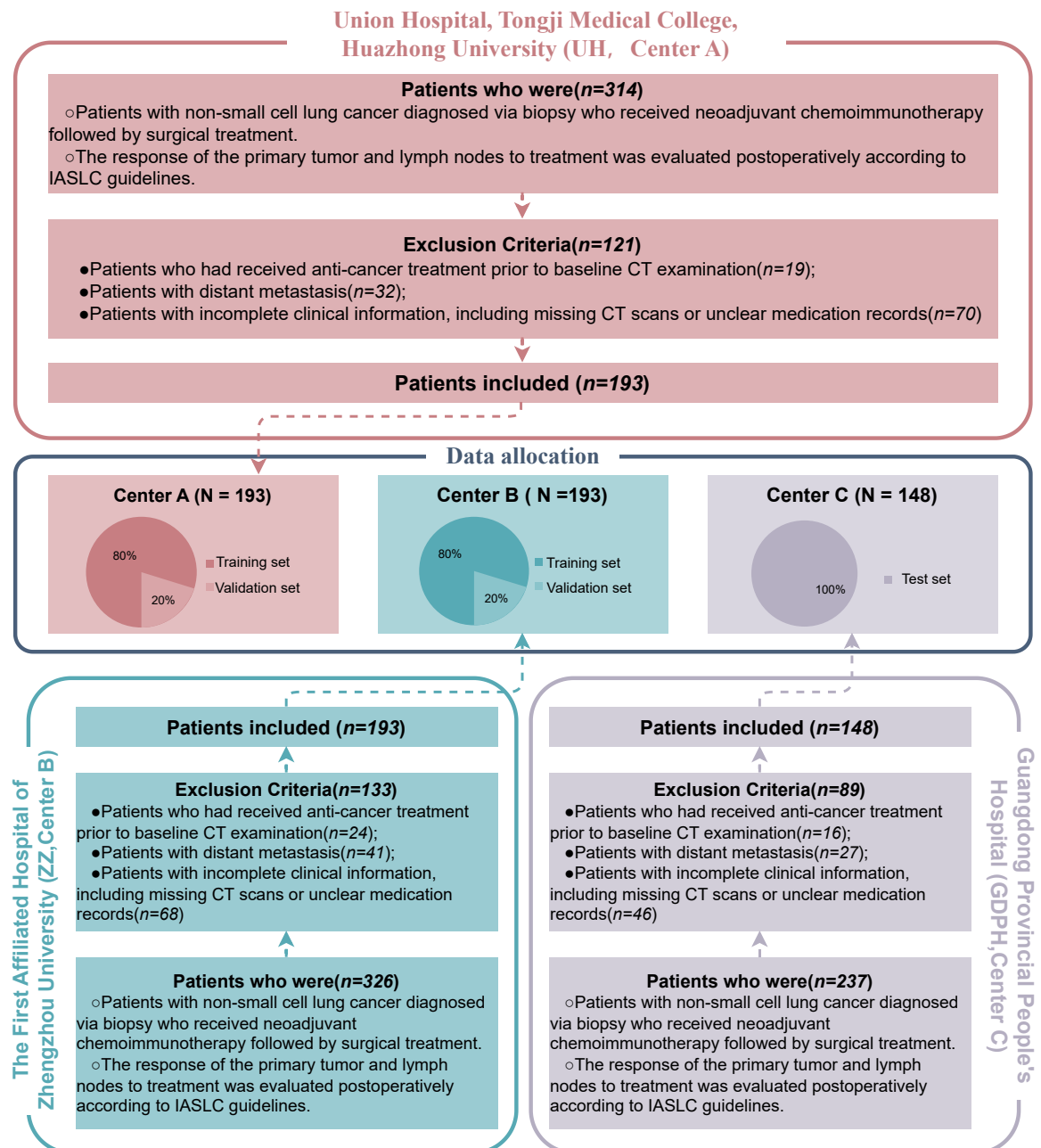


Fig. 1: Flowchart shows patient exclusion for each center.

to the internal validation dataset, following an 8:2 split ratio. The 148 patients from Center C formed an independent and external test dataset, which was used for model evaluation. Clinical information, blood indices, and CT image data before and after treatment were collected, more details see [Supplementary Methods](#) (Data Acquisition). To ensure patient confidentiality, all imaging data were anonymized by removing or encrypting identifiable information from the Digital Imaging and Communications in Medicine (DICOM) headers, including patient names, IDs, dates of birth, and institutional details, using a standardized anonymization tool compliant with the DICOM standard. We conducted the study following Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + artificial intelligence (TRIPOD + AI) guidelines.³¹

Definition of pathologic complete response and clinical complete response

Pathologic response was assessed according to International Association for the Study of Lung Cancer (IASLC) guidelines by experienced pathologists evaluating the primary tumor and lymph nodes in patients with NSCLC following NACI.³² To ensure accuracy, the initial evaluation was performed by the primary physician and subsequently reviewed by a senior pathologist. pCR was defined as the absence of residual viable tumor cells (ypT0 and ypN0) in both the tumor bed and lymph nodes. A radiologist with 10 years of experience in lung cancer diagnosis (G. Wu), evaluates the lesions according to the RECIST 1.1 criteria,³³ which are then reviewed by a senior radiologist (J. Gao). The Complete Response (CR) is defined as the complete disappearance of all target lesions, as measured by the longest diameter, with no evidence of new lesions. Additionally, all non-target lesions must be fully resolved (e.g., lymph nodes reduced to normal size).

Image acquisition

All patients underwent contrast-enhanced CT scans at two time points: a baseline scan (T_0) acquired within 14 days prior to treatment initiation, and a pre-surgical scan (T_1) obtained 7–14 days before planned resection. The LC-NICER_α model utilized T_0 scans exclusively for pretreatment response prediction, whereas the LC-NICER_β model integrated both T_0 and T_1 scans to guide surgical decision-making. More details see [Supplementary Methods](#) (Image acquisition and preprocessing).

Multi-dimensional feature extraction and selection

The feature extraction process, using a multi-dimension and multi-scale approach, integrated radiomics, deep learning, and habitat imaging techniques to comprehensively analyze the tumor and its microenvironment. More details see [Supplementary Methods](#). CT

images were utilized as the primary data source, and features were extracted to quantify tumor heterogeneity, spatial relationships, and microenvironmental characteristics.

To ensure the robustness and relevance of the extracted features, we implemented a comprehensive feature selection pipeline applied uniformly to radiomics, deep learning, and habitat imaging features. First, features with low variance (<75%) were removed to retain only those with sufficient variability to discriminate between samples. Subsequently, the Spearman correlation coefficient was calculated to eliminate highly correlated features (>0.75), reducing redundancy and multicollinearity. Then, a t-test or Mann–Whitney U test was performed to compare feature values between positive and negative samples, retaining only those with a *p*-value less than 0.05. Finally, Least Absolute Shrinkage and Selection Operator regression with 5-fold cross-validation was applied to automatically select the most informative features, ensuring a sparse set of key features while minimizing overfitting. This rigorous pipeline ensured that only reproducible, discriminative, and non-redundant features were included in the final models.

Predictive modeling architecture

The predictive modeling process was designed to integrate multi-dimensional and multi-temporal features for predicting pCR in patients with NSCLC undergoing NACI. The architecture was structured into three key components, forming a hierarchical framework. (1) Base Model Development: Individual models were built for radiomics, deep learning, and habitat imaging features across three temporal dimensions (pre-treatment, post-treatment, and delta changes) using MediAI Hub, generating respective radiomics-score, deep learning (DL)-score, and habitat imaging (HI)-score. (2) Multi-dimensional Integration: These feature scores were combined to create comprehensive models for each temporal dimension, yielding PreScore, PostScore, and DeltaScore. (3) Temporal Integration: The final predictive model integrated these temporal scores via MediAI Hub, leveraging the complementary contributions of multi-modal features and temporal dynamics. Detailed methodology is provided in [Supplementary Methods](#) (Model Construction and Evaluation).

Quantitative evaluation and statistical analysis

Model performance was evaluated using the area under the curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV), with 95% confidence intervals calculated via bootstrapping (1000 iterations) to ensure reliability. Statistical significance of performance differences was assessed using the Wilcoxon signed-rank test or paired t-test (as appropriate) for overall metrics, and the DeLong test for AUC

comparisons, with a significance threshold of $p < 0.05$. Categorical variables were analyzed using Fisher's exact test or Pearson's chi-square test as appropriate, while continuous variables were assessed through Student's t -test or Mann–Whitney U test based on distributional assumptions. For subgroup analysis, Benjamini–Hochberg correction approach was used for multiple comparisons. Statistical significance was defined as a two-sided p -value < 0.05 using a predefined threshold. To assess generalizability, stratified analyses were conducted across key subgroups, such as tumor stage and histology, ensuring consistent performance across diverse patient populations. Shapley additive explanations analysis was performed to quantify the impact of each temporal dimension score on the model prediction. All computations and statistical analyses were performed using Python (version 3.9; Python Software Foundation), and R software (version 3.6.0; <https://www.r-project.org>).

Role of the funding source

The funders of this study had no roles in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Participants baseline information

Patients were classified into pCR and non-pCR groups based on the presence of residual tumor in the post-operative specimens. We compared the clinicopathological features between the pCR and non-pCR groups in [Table 1](#). Statistical analysis revealed significant intergroup differences in demographic and clinical parameters, including gender distribution, smoking history, pathological classification, T-stage status, pretherapeutic lactic dehydrogenase (Pre-LDH) levels, post-therapeutic LDH (Post-LDH) values, along with post-treatment systemic immune-inflammation index

Characteristics	Total (n = 534)	Non-pCR (n = 332)	pCR (n = 202)	p value
Age (years), Mean $\pm$ SD	59.974 $\pm$ 8.452	59.742 $\pm$ 8.703	60.334 $\pm$ 8.031	0.436
Gender, n (%)				0.001 ^a
Female	75 (14.0)	59 (17.8)	16 (7.9)	
Male	459 (86.0)	273 (82.2)	186 (92.1)	
Smoking, n (%)				0.002 ^a
Yes	311 (58.2)	176 (53.0)	135 (66.8)	
No	223 (41.8)	156 (47.0)	67 (33.2)	
Pathological type, n (%)				<0.001 ^a
Squamous carcinoma	368 (68.9)	207 (62.3)	161 (79.7)	
Adenocarcinoma	133 (24.9)	107 (32.2)	26 (12.9)	
Other	33 (6.2)	18 (5.4)	15 (7.4)	
T stage, n (%)				0.005 ^a
T1	68 (12.7)	33 (9.9)	35 (17.3)	
T2	211 (39.5)	124 (37.3)	87 (43.1)	
T3	153 (28.7)	100 (30.1)	53 (26.2)	
T4	102 (19.1)	75 (22.7)	27 (13.4)	
N stage, n (%)				0.715
N0	46 (8.6)	31 (9.3)	15 (7.4)	
N1	131 (24.5)	82 (24.7)	49 (24.3)	
N2	324 (60.7)	201 (60.5)	123 (60.9)	
N3	33 (6.2)	18 (5.4)	15 (7.4)	
Clinical stage, n (%)				0.329
II	115 (21.5)	67 (20.2)	48 (23.8)	
III	419 (78.5)	265 (79.8)	154 (76.2)	
Pre LDH (IU/L), median, [IQR]	184.00 [172.000–213.250]	184.000 [169.250–215.750]	192.500 [174.000–209.250]	0.049 ^a
Pre Albumin (g/dL), median, [IQR]	39.650 [36.700–42.200]	39.700 [37.000–42.200]	39.600 [35.725–42.225]	0.370
Pre SII, median, [IQR]	699.190 [460.360–1070.770]	699.370 [464.550–1081.650]	695.082 [455.819–1061.724]	0.960
Pre NLR, median, [IQR]	2.840 [2.040–3.890]	2.870 [2.060–3.880]	2.780 [2.001–3.917]	0.738
Post LDH (IU/L), median, [IQR]	197.000 [178.750–211.000]	197.000 [180.000–212.750]	196.000 [177.000–209.250]	0.037 ^a
post Albumin (g/dL), median, [IQR]	39.800 [37.675–42.725]	39.800 [37.625–42.600]	40.000 [37.675–42.800]	0.505
Post SII, median, [IQR]	412.898 [263.380–661.434]	449.838 [280.008–696.026]	373.544 [248.585–593.595]	0.020 ^a
Post NLR, median, [IQR]	2.113 [1.520–3.007]	2.262 [1.618–3.181]	1.952 [1.345–2.796]	0.005 ^a

Abbreviations: SD, standard deviation; pCR, pathological complete response; LDH, lactic dehydrogenase; IQR, interquartile range; NLR, neutrophil to lymphocyte ratio; SII, systemic immune-inflammation index. ^aIndicates statistical significance: p value < 0.05 .

Table 1: Comparison of clinical and pathological characteristics between pCR group and non-pCR group.

(Post-SII) and neutrophil-to-lymphocyte ratio (Post-NLR). Table 2 presents the characteristics of patients across the three cohorts. Besides, we have conducted a thorough analysis of potential differences in patient populations and model performance among the three centers. Regarding the differences in baseline characteristics of patients, a detailed comparison was made among the baseline characteristics of three centers (A, B, C). As expected, there are some inherent inter center differences (e.g., slightly lower proportion of squamous cell carcinoma and more stage II patients in center C). Overview of the study design and analytical workflow is shown in Fig. 2.

Model performance for treatment response prediction

As illustrated in Fig. 3, we developed predictive models for treatment response to NACI using longitudinal imaging data and multi-dimensional features. The performance of these models, evaluated by the AUCs at different time points across various datasets, are shown in Fig. 3A. The LC-NICER_δ, which combines features extracted from pre-treatment CT, post-treatment CT, and delta changes through radiomics, deep learning,

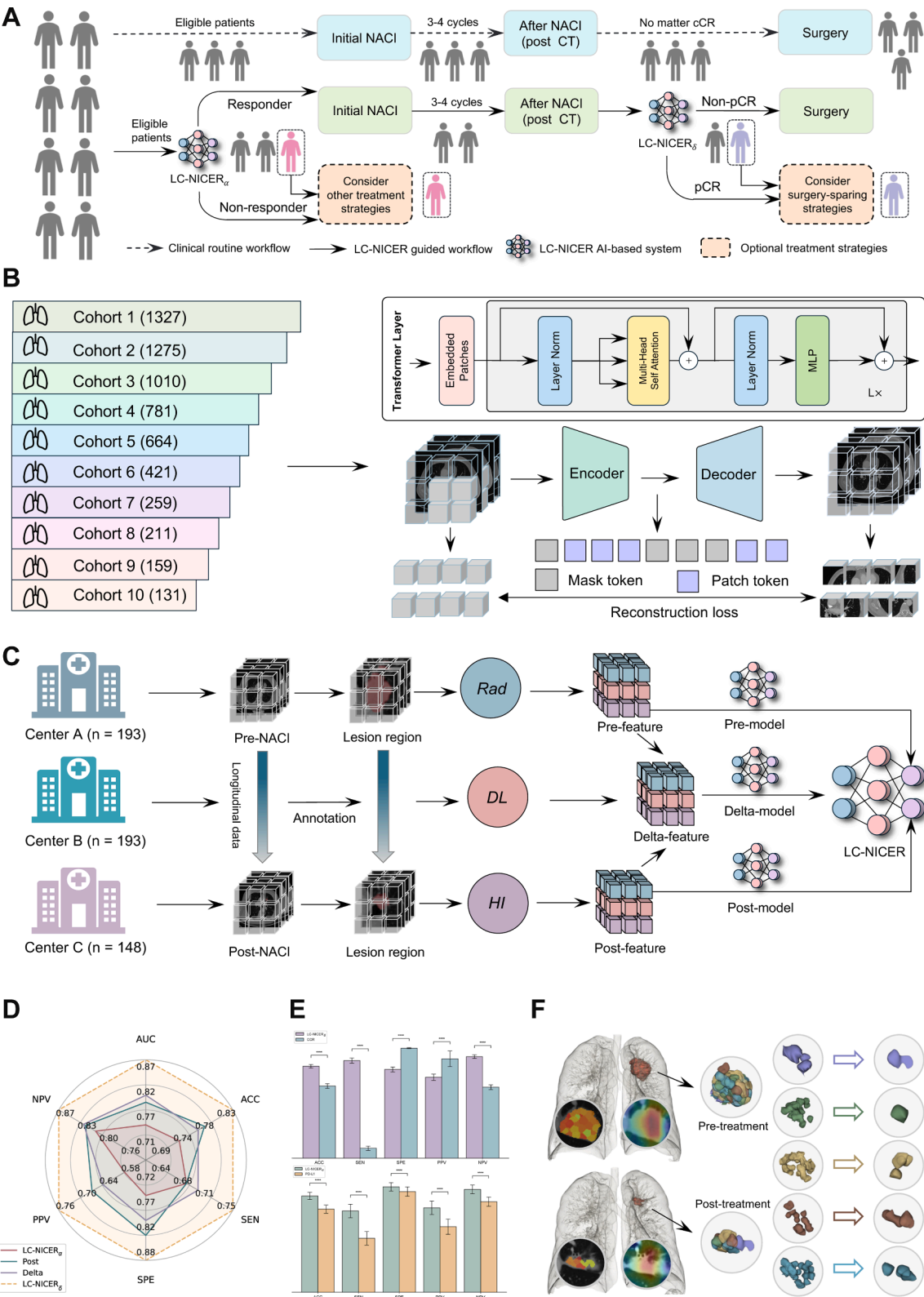
and HI analysis, demonstrated robust performance with AUCs of 0.950, 0.889, and 0.870 in the training, internal validation, and external test datasets, respectively. Furthermore, in the test dataset, the model utilizing delta features demonstrated superior performance, achieving an AUC of 0.796 compared to the pre-treatment (AUC: 0.734) and post-treatment (AUC: 0.781) models. Additionally, across the pre-, post-, and delta-feature categories, radiomics and habitat imaging features were found to play dominant roles in predictive performance.

The overall performance metrics of the models, including AUC, ACC, SEN, SPE, PPV, and NPV, are summarized in Fig. 3B. The LC-NICER_δ outperformed others across all metrics. To further analyze feature contributions, Kruskal–Wallis tests were conducted (Fig. 3C), revealing statistically significant differences ($p < 0.001$) among radiomics, deep learning, and habitat imaging scores, indicating they capture distinct aspects of tumor characteristics. Prior to NACI treatment initiation, LC-NICER_α surpassed the PD-L1 biomarker (Fig. 3D), achieving an ACC of 0.722 (95% CI: 0.668–0.772) compared to PD-L1's 0.622 (95% CI: 0.564–0.683), with higher SEN (0.609 vs. 0.404) and

Characteristics	Total (n = 534)	Training (n = 308)	Validation (n = 78)	Test (n = 148)	p value
Age (years), Mean ± SD	59.973 ± 8.452	60.182 ± 8.278	61.602 ± 7.393	58.664 ± 9.167	0.168
Gender, n (%)					0.332
Female	75 (14.0)	38 (12.3)	11 (14.1)	26 (17.6)	
Male	459 (86.0)	270 (87.7)	67 (85.9)	122 (82.4)	
Smoking, n (%)					0.637
Yes	311 (58.2)	176 (57.1)	44 (56.4)	91 (61.5)	
No	223 (41.8)	132 (42.9)	34 (43.6)	57 (38.5)	
Pathological type, n (%)					0.002 ^a
Squamous carcinoma	368 (68.9)	222 (72.1)	62 (79.5)	84 (56.8)	
Adenocarcinoma	133 (24.9)	70 (22.7)	14 (17.9)	49 (33.1)	
Other	33 (6.2)	16 (5.2)	2 (2.6)	15 (10.1)	
Clinical T stage, n (%)					0.259
T1	68 (12.7)	31 (10.1)	13 (16.7)	24 (12.7)	
T2	211 (39.5)	121 (39.3)	26 (33.3)	64 (39.5)	
T3	153 (28.7)	95 (30.8)	23 (29.5)	35 (28.7)	
T4	102 (19.1)	61 (19.8)	16 (20.5)	25 (19.1)	
Clinical N stage, n (%)					<0.001 ^a
N0	46 (8.6)	22 (7.1)	5 (6.4)	19 (12.8)	
N1	131 (24.5)	64 (20.8)	22 (28.2)	45 (30.4)	
N2	324 (60.7)	205 (66.6)	50 (64.1)	69 (46.6)	
N3	33 (6.2)	17 (5.5)	1 (1.3)	15 (10.2)	
Clinical stage, n (%)					0.007 ^a
II	115 (21.5)	58 (18.8)	12 (15.4)	45 (30.4)	
III	419 (78.5)	250 (81.2)	66 (84.6)	103 (69.6)	
pCR, n (%)					0.727
Present	202 (37.8)	120 (39.0)	30 (38.5)	52 (35.1)	
Absent	332 (62.2)	188 (61.0)	48 (61.5)	96 (64.9)	

Abbreviations: SD, standard deviation; pCR, pathological complete response. ^aIndicates statistical significance: p value < 0.05.

Table 2: Patient characteristics across different cohorts.



SPE (0.790 vs. 0.754). Following NACI treatment completion, LC-NICER_s significantly outperformed the RECIST criteria (Clinical Complete Response, CCR) in assessing complete response (Fig. 3E, details in Table 3). LC-NICER_s achieved a higher ACC of 0.831 (95% CI: 0.800–0.861) than CCR's 0.651 (95% CI: 0.610–0.693), with markedly higher SEN (0.881 vs. 0.089) and comparable SPE (0.802 vs. 0.994). Statistical analysis using the Wilcoxon test confirmed significant differences ($p < 0.05$), highlighting the superior accuracy and efficiency of our AI-driven system. Fig. 3F displays the Spearman correlation coefficients among radiomics, deep learning, HI features, where the HI category involved HI tumor and HI peritumors (i.e., 3 mm, 5 mm and 7 mm). While radiomics, deep learning, and HI features exhibited low inter-correlation, moderate to high correlations were observed within the HI peritumor category (HI Peri-3, HI Peri-5, and HI Peri-7), likely due to overlapping regions in the peritumoral areas.

Fig. 4 comprehensively evaluates the performance and interpretability of the LC-NICER_s. In Fig. 4a, the model's performance metrics are presented across the training, validation, and test datasets, demonstrating robust performance and generalization without significant statistical differences ($p > 0.05$) among the three datasets. Fig. 4b displays the confusion matrices for the three datasets, with the cut-off value set at 0.27 based on the Youden index, further validating the model's consistency in classifying treatment response. Fig. 4c illustrates the t-distributed stochastic neighbor embedding (t-SNE) visualization of the feature space, showing clear separation between pCR and non-pCR groups, highlighting the discriminative power of the extracted features. Fig. 4d presents the SHapley Additive exPlanations (SHAP) values, revealing that the post-treatment score consistently had the highest average impact on the model's output magnitude across all datasets. This observation aligns with clinical practice, where the presence or absence of a well-defined lesion region in post-treatment CT images serves as a key indicator for clinicians to assess pCR. Fig. 4e shows the subgroup analysis of the LC-NICER_s, split by age, sex, T stage, clinical stage, histopathological type, pre SII, post SII, pre NLR, post NLR, and post LDH across all datasets.

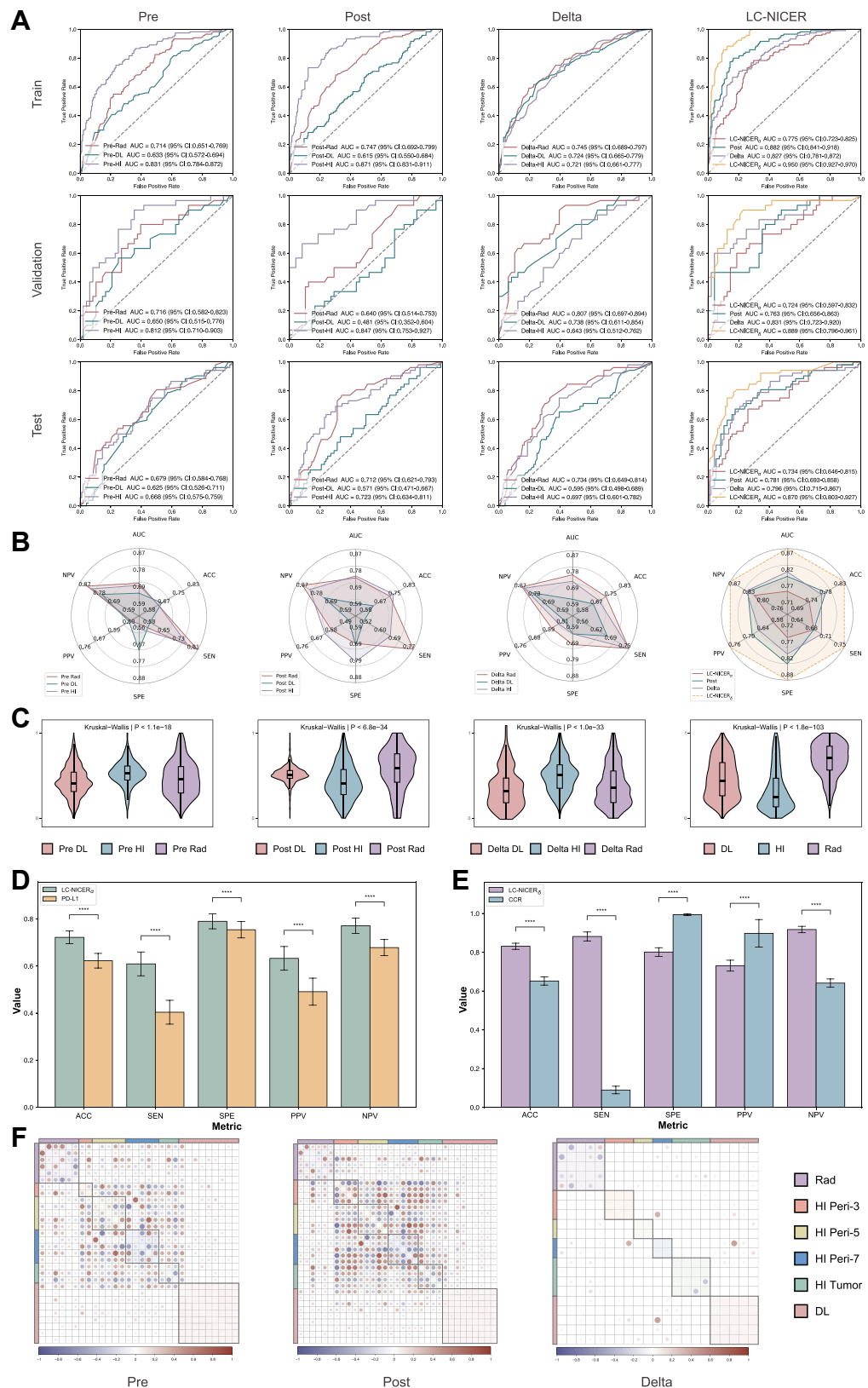
Habitat imaging analysis between pre-treatment and post-treatment

Habitat imaging analysis revealed dynamic spatiotemporal relationship (DSR) changes within subregions (SRs) between pre- and post-treatment CT scans, which we categorized into five distinct patterns: (1) DSR appearance (newly emerged SRs), (2) DSR disappearance (vanished SRs), (3) DSR increase (enhanced feature values), (4) DSR decrease (reduced feature values), and (5) DSR unchanged (SRs not changed significantly). Through rigorous feature selection, we identified key subregions significantly associated with pCR prediction. Importantly, DSR changes were quantified based on habitat imaging feature dynamics (e.g., texture or intensity patterns) rather than spatial volume alterations, with a 5% change threshold implemented to account for inherent variability in the measurements.

Fig. 5A–D depict the habitat analysis results for the tumor and peritumor regions at 3 mm, 5 mm, and 7 mm distances, respectively. These visualizations highlight the spatial heterogeneity and temporal evolution of SRs, providing insights into how different SRs respond to NACI. Note that, the SRs before and after treatment do not correspond to spatially identical regions but rather to the same SR categories and their characteristic attributes. Fig. 5E presents a statistical bar plot summarizing the distribution of five subregion change types across all analyzed tumor and peritumoral subregions, with bars within each region representing serial measurements taken before and after treatment. This analysis, based on data from all study cohorts, underscores the importance of habitat imaging in capturing tumor heterogeneity and identifies specific subregion changes strongly associated with treatment response, providing potential biomarkers for personalized therapeutic strategies.

Quantitative radiomic analysis of metabolic activity patterns, as measured by energy (representing signal uniformity), entropy (quantifying tissue heterogeneity) and mean intensity (reflecting perfusion), revealed distinct spatial distributions of treatment-responsive subregions (Fig. 5F–H). Although global tumor measurements showed no significant intergroup differences (pCR vs non-pCR, $p > 0.05$), habitat mapping identified spatially conserved metabolic niches with

Fig. 2: Overview of the study design and analytical workflow. **A.** Clinical routine workflow and LC-NICER guided workflow. **B.** CT foundation model pre-training through a large-scale and diverse cohort. **C.** Multi-dimensional feature extraction and longitudinal model construction. A foundation model was employed to extract high-resolution features from longitudinal CT data. Delta-features were derived to capture changes between pre- and post-treatment states, enabling the construction of a longitudinal model for lesion analysis. **D.** Quantitative evaluation and statistical analysis: Model performance was assessed using metrics. **E.** Comparison of traditional indicators and statistical testing. **F.** Visualization was performed to interpret the results and validate the clinical utility of the model. Abbreviations: LC-NICER, Lung Cancer Neo-adjuvant Immuno-ChEmotherapy Response predictor; NACI, Neoadjuvant chemioimmunotherapy; NSCLC, non-small cell lung cancer; pCR, pathological complete response; Rad, radiomics; DL, deep learning; HI, habitat imaging; CCR, Clinical Complete Response.



Method	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPV (95% CI)	NPV (95% CI)	p value ^a
LC-NICER ₈	0.831 (0.800–0.861)	0.881 (0.831–0.927)	0.802 (0.758–0.844)	0.730 (0.671–0.782)	0.916 (0.883–0.946)	$p < 0.001^b$
CCR	0.651 (0.610–0.693)	0.089 (0.052–0.134)	0.994 (0.984–1.000)	0.905 (0.750–1.000)	0.642 (0.597–0.684)	
LC-NICER _α	0.722 (0.668–0.772)	0.609 (0.509–0.705)	0.790 (0.728–0.848)	0.632 (0.535–0.728)	0.771 (0.706–0.837)	$p < 0.001^b$
PD-L1	0.622 (0.564–0.683)	0.404 (0.301–0.505)	0.754 (0.686–0.822)	0.491 (0.378–0.600)	0.678 (0.610–0.740)	

Abbreviations: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; PPV, Positive Predictive Value; NPV, Negative Predictive Value; CCR, Clinical Complete Response; CI, Confidence Interval. ^aWilcoxon test. ^bIndicates statistical significance: p value < 0.05 .

Table 3: Comparison of LC-NICER_α and LC-NICER₈ with clinical methods.

differential response characteristics: metabolically uniform regions (SR1, SR2 and SR4 in Fig. 5F), areas of high metabolic disorder (SR1, SR3 in Fig. 5G) and high-perfusion zones (SR2, SR4, SR5 in Fig. 5H). Notably, SRs exhibiting significant metabolic activation differences between pCR and non-pCR groups demonstrated higher odds ratios (OR) compared to those with non-significant variations. This spatial mapping of metabolic microenvironments provides a biological framework for understanding regional treatment susceptibility and resistance patterns, offering clinically actionable insights for targeted therapy optimization.

Spatiotemporal distribution of key subregions

Fig. 6 comprehensively explores the importance and distribution of key SRs in predicting pCR to NACI treatment. The importance of these key SRs was further quantified using SHAP values, which revealed their relative contributions to the model's predictive performance. Fig. 6A and B present box plots summarizing the importance of 20 original SRs (5 from the tumor and 15 from peritumor regions) at pre-treatment and post-treatment time points, stratified by pCR and non-pCR groups. These plots reveal dynamic changes in SR importance over time, highlighting which SRs are more influential in predicting treatment response.

Fig. 6C and D further illustrate the distribution of key SRs using pie charts. Each pie chart shows the percentage of pCR and non-pCR cases for a specific SR, with Fig. 6C representing pre-treatment data and Fig. 6D representing post-treatment data. Our analysis identified several SRs with significant predictive value: Post_peri5_SR4 and post_peri7_SR4 were exclusively associated with non-pCR (100% and 71.4% specificity)

when they have the greatest importance in cases. These SRs typically exhibited irregular contours, heterogeneous enhancement, scattered distribution, and spiculated margins, which are imaging features associated with treatment resistance. In contrast, Post_peri5_SR1 and Pre_tumor_SR3 showed strong associations with pCR, with probabilities of 83.3% and 50.4%, respectively. These SRs were characterized by well-defined, smooth contours, homogeneous enhancement, and the absence of spiculated margins, reflecting favorable treatment responses. Fig. 6E–H shows four typical cases with imaging interpretation of these four key SRs. These findings underscore the spatial and temporal heterogeneity of tumor response to therapy and provide potential clinical guidance. For example, the presence of specific SRs could help identify patients who may benefit from alternative therapies or avoid unnecessary surgery, thereby optimizing treatment strategies.

Clinical application and case studies

Fig. 7A demonstrates the clinical applicability of the LC-NICER_α and LC-NICER₈ in predicting therapy response at both early and late stages of treatment. In Fig. 7B, a 63-year-old male patient with stage IIIA squamous cell carcinoma, a tumor diameter of 4.0 cm, and low PD-L1 expression (Tumor cell Proportion Score [TPS]: 5%) was classified as non-CCR by RECIST. Our LC-NICER_α predicted a pCR probability of 83.1% at the early stage, while LC-NICER₈ increased this probability to 90.2% at the late stage. Post-surgical pathology confirmed pCR, validating the accuracy of our models. Conversely, Fig. 7C presents a 64-year-old male patient with similar clinical characteristics (stage IIIB, tumor diameter of 3.1 cm, TPS: 70%) who was classified as CCR by

Fig. 3: Performance evaluation of the predictive models. **A.** The AUCs at different time points across training, internal validation, and external test datasets. Performance was evaluated for various imaging feature types, including Rad, DL, and HI features, as well as temporal types (pre-/post-treatment, and delta features). The final integrated model, LC-NICER₈, was also assessed. **B.** Radar charts illustrating the overall performance metrics for the different models. **C.** Statistical analysis using Kruskal-Wallis tests to compare the radiomics score, deep learning score, and HI score across pre-, post-treatment, and delta features, demonstrating that the three feature types capture distinct and complementary information. **D.** Bar charts showing comparison of the overall performance between LC-NICER_α and PD-L1. **E.** Bar charts showing comparison of the overall performance between LC-NICER₈ and CCR. **F.** Spearman correlation coefficients were calculated to assess the relationships between radiomics, deep learning, and HI features within pre-treatment, post-treatment, and delta feature sets, further highlighting their unique contributions to the model. Abbreviations: AUCs, Area under the curves; Rad, radiomics; DL, deep learning; HI, habitat imaging; CCR, Clinical Complete Response.

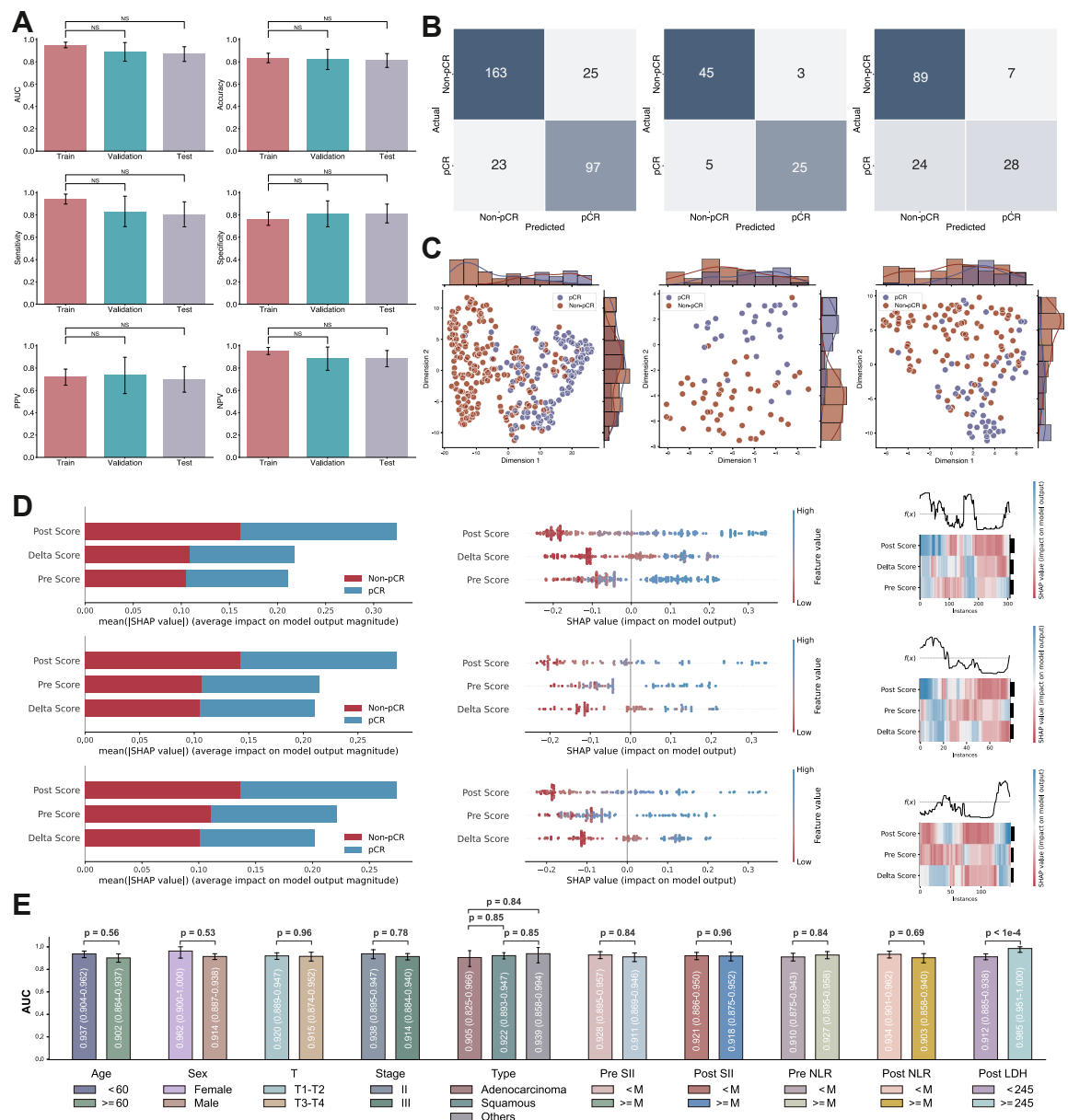
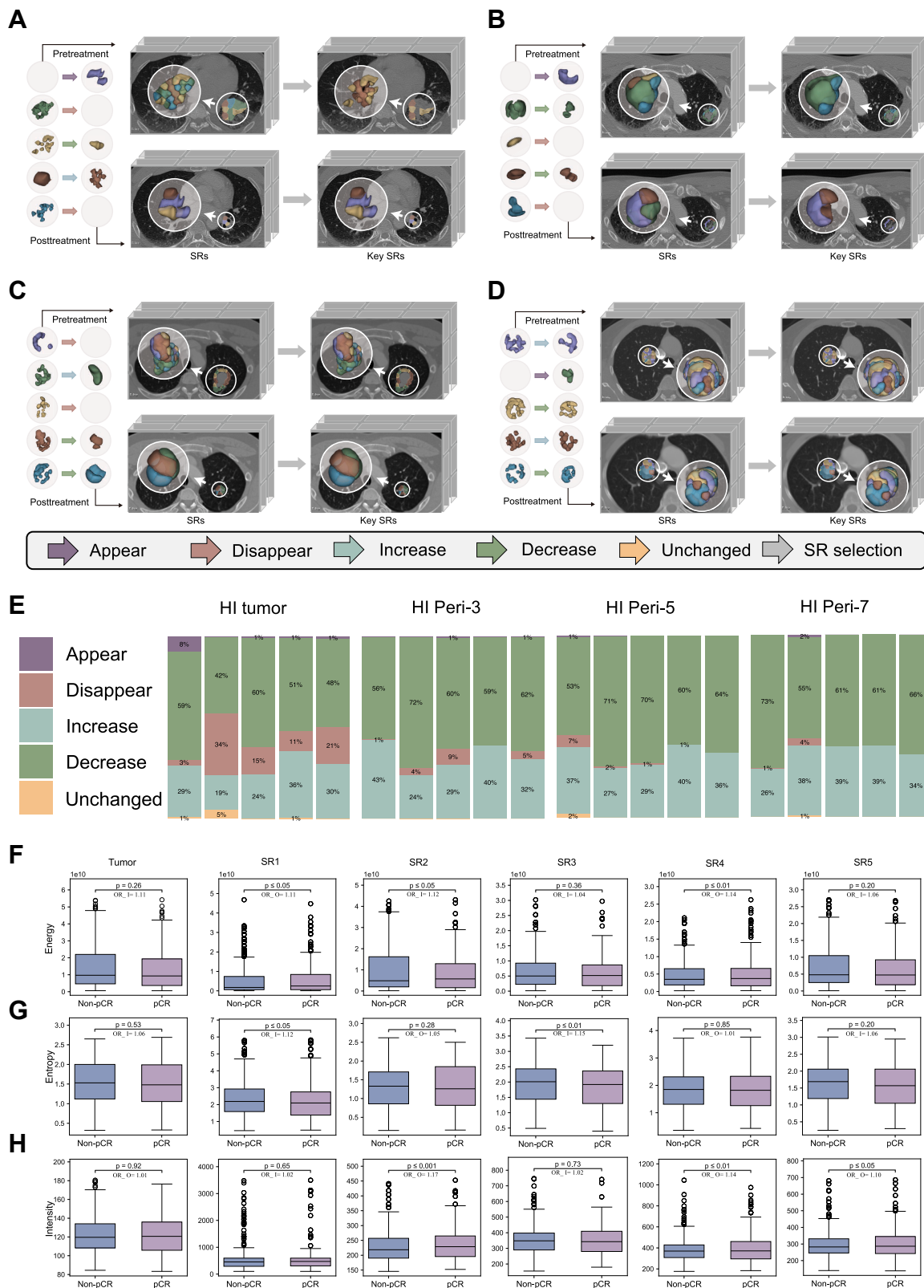


Fig. 4: The performance and interpretability of the LC-NICER. **a.** Bar plots showing the model's performance metrics (AUC, ACC, SEN, SPE, NPV, and PPV) across the training, validation, and test datasets. Wilcoxon test was performed to evaluate the differences in performance among the three datasets. **b.** Confusion matrices for the three datasets, illustrating the classification performance of the LC-NICER₈ in distinguishing between different clinical outcomes. **c.** t-SNE visualization of the feature space, highlighting the separation and clustering of samples based on radiomics, deep learning, and HI features. **d.** SHAP values for the LC-NICER₈ demonstrating the contribution of individual features to the model's predictions and providing insights into the interpretability of the model. **e.** Subgroup analysis of the LC-NICER₈ across all datasets. Abbreviations: AUC, the area under the curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; t-SNE, the t-distributed stochastic neighbor embedding; SHAP, the SHapley Additive exPlanations.

RECIST. Our LC-NICER_α and LC-NICER₈ predicted non-pCR probabilities of 81.1% and 87.5%, respectively, which were later confirmed by post-surgical pathology.

Beyond prediction accuracy, our AI-driven system provides comprehensive visualizations to enhance clinical interpretability. These include subregion

visualizations of the tumor and peritumoral regions (2D and 3D), highlighting key SRs identified through their predictive importance. Additionally, feature heatmaps generated from radiomics and deep learning analyses are provided, with highlighted areas indicating suspicious non-pCR regions. These visualizations offer



clinicians actionable insights into tumor heterogeneity and treatment response dynamics, supporting informed decision-making. For easy clinical utility and research reproducibility, we developed and openly published a software (i.e., LC-NICER). More details see [Supplementary Fig. S1](#).

Discussion

There is an urgent need for universally applicable predictive biomarkers to identify responders and assess therapeutic efficacy in patients with NSCLC receiving NACI.³⁴ Current biomarkers like PD-L1 and RECIST face limitations, including sampling bias, temporal heterogeneity, lack of standardization, and inability to fully capture immune response dynamics.³⁵ To address these challenges, we propose LC-NICER, a hierarchical framework integrating multi-dimensional and longitudinal data. LC-NICER combines radiomics, deep learning, and habitat imaging scores, enabling a comprehensive and non-invasive approach to predict treatment response and guide precision immunotherapy in lung cancer. Specifically, it can offer preliminary risk stratification for patients awaiting histological results, identify subtle imaging patterns that may warrant closer pathological scrutiny, and assist in monitoring treatment response when repeated biopsies are clinically impractical. This synergy leverages the strengths of both AI-driven imaging analysis (rapid, non-invasive, whole-lesion assessment) and histological evaluation (gold-standard tissue characterization).

LC-NICER demonstrated significant advantages over FDA-approved biomarkers (e.g., PD-L1) and RECIST in identifying immunotherapy-responsive populations and predicting treatment response. PD-L1 expression, while widely used, is limited by its spatial heterogeneity and the invasive nature of tissue sampling,³⁶ which may result in incomplete or inaccurate assessments of tumor biology. In contrast, our non-invasive CT-based approach provides a comprehensive evaluation of both the tumor and its surrounding microenvironment, offering a more reliable and holistic assessment. Similarly, RECIST often misclassify patients with pCR as SD or PR due to pseudo-lesions caused by immune cell infiltration.³⁷ This limitation may result in unnecessary surgical interventions and suboptimal treatment decisions. The proposed LC-NICER₆ overcomes these challenges, achieving an accuracy of 0.831 (95% CI:

0.800–0.861), significantly higher than CCR's 0.651 (95% CI: 0.610–0.693). Moreover, LC-NICER_α outperformed PD-L1 with an accuracy of 0.722 (95% CI: 0.668–0.772) compared to PD-L1's 0.622 (95% CI: 0.564–0.683), demonstrating superior predictive performance. These results highlight the limitations of current clinical metrics, which are constrained by their reliance on invasive sampling, spatial heterogeneity, and susceptibility to imaging artifacts. In contrast, our AI-driven framework leverages multi-dimensional imaging features to provide accurate, non-invasive, and cost-effective predictions, enhancing clinical decision-making and potentially reducing unnecessary interventions. By integrating spatiotemporal heterogeneity and advanced AI techniques, LC-NICER offer a transformative approach to optimizing NACI for NSCLC.

Previous studies predominantly relied on single-dimensional features, such as radiomics³⁸ or deep learning,³⁹ or data from isolated time points for predicting therapy response.⁴⁰ Lin et al. developed a radiomics-based model for predicting pCR in NSCLC by extracting a single radiomics feature from pre-neoadjuvant immunotherapy CT imaging data.⁴¹ Qu et al. utilized CT data prior to neoadjuvant immunotherapy for NSCLC to extract deep learning features and develop a predictive model for pCR.⁴² Ye et al. established a predictive model for pCR by analyzing habitat features derived from the same type of CT imaging data prior to neoadjuvant immunotherapy in NSCLC patients.⁴³ These approaches often fail to capture the dynamic and heterogeneous nature of tumor and peritumor response to therapy. To address this limitation, we first extracted HI features to comprehensively quantify intra- and peritumoral heterogeneity. We then integrated radiomics, deep learning and HI features to construct a multi-dimensional feature fusion model, significantly enhancing the predictive accuracy of treatment response compared to the single-dimensional feature models.

Prior research has demonstrated that leveraging multiple imaging datasets during treatment can enhance the model's ability to predict treatment efficacy.⁴⁴ Jin et al. employed MRI data both before and after neoadjuvant chemotherapy for rectal cancer to derive deep learning features for constructing a prediction model for pCR, which exhibited enhanced predictive performance in comparison to models solely

Fig. 5: Habitat imaging analysis of tumor and peritumoral regions before and after NACI treatment. **A.** Habitat imaging analysis of the tumor core. **B–D.** Habitat imaging analysis of peritumoral regions at 3 mm (B), 5 mm (C), and 7 mm (D) distances from the tumor boundary, comparing pre-treatment and post-treatment states. **E.** Distribution of five dynamic spatiotemporal changes (appear, disappear, increase, decrease, unchanged) across tumor and peritumoral subregions. Bars show pre- and post-treatment serial measurements, highlighting habitat dynamics post-treatment. Results from all study cohorts. **F.** Energy distribution (metabolic uniformity) highlights organized metabolic niches (SR1, 2, 4). **G.** Entropy analysis (metabolic disorder) reveals chaotic growth patterns (SR1, 3). **H.** Mean intensity mapping (perfusion surrogate) identifies vascularized subregions (SR2, 4, 5). Abbreviations: NACI, Neoadjuvant chemioimmunotherapy; HI, habitat imaging; SR, subregion; OR₀, original odds ratio; OR_I, inverted odds ratio.

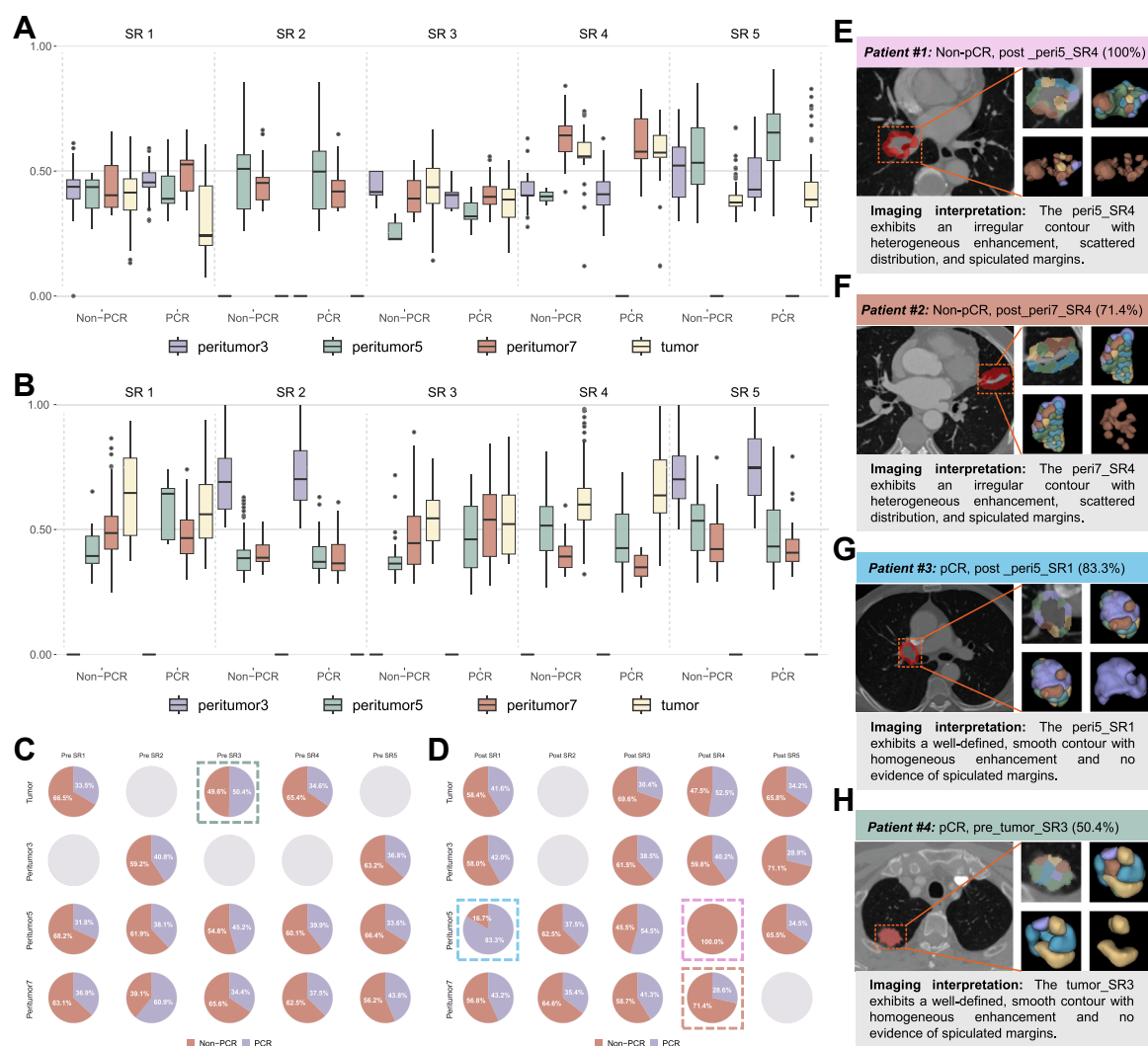


Fig. 6: Importance and distribution of key SRs in predicting pCR. **A.** Importance of the 20 original SRs at pre-treatment. **B.** Importance of the 20 original SRs at post-treatment. **C.** Distribution of key SRs at pre-treatment, represented in a 4 × 5 grid (rows: tumor, peri3, peri5, peri7; columns: SR1–SR5). **D.** Distribution of key SRs at post-treatment, organized similarly to **C**. **E–H.** Representative SRs with significant predictive value. **E.** Patient#1 (non-pCR): Post_peri5_SR4 (100%) showing irregular contour, heterogeneous enhancement, scattered distribution, and spiculated margins. **F.** Patient#2 (non-pCR): Post_peri7_SR4 (71.4%) exhibiting irregular, lobulated contour with heterogeneous enhancement and spiculated margins. **G.** Patient#3 (pCR): Post_peri5_SR1 (83.3%) demonstrating a well-defined, smooth contour with homogeneous enhancement and no spiculated margins. **H.** Patient#4 (pCR): Pre_tumor_SR3 (50.4%) showing a well-defined, smooth contour with homogeneous enhancement and no spiculated margins. Abbreviations: SRs, subregions; pCR, pathological complete response.

based on pre-treatment data.⁴⁵ In our analysis, the Pre-Model, Post-Model, and Delta-Model achieved AUC values of 0.734, 0.781, and 0.796, respectively, in the test set. These outcomes suggest that dynamic models outperform both the pre-treatment and post-treatment models that rely on data from single time points. Additionally, by leveraging longitudinal CT imaging data acquired before and after NACI for NSCLC (temporal dimension) and combining it with multi-dimensional spatial tumor characterization, we developed a spatiotemporal fusion model for pCR prediction, achieving an AUC value of 0.870 in the test set. This

innovative model provides a robust validated framework to guide clinical decisions, potentially advancing personalized treatment strategies in NSCLC.

Emerging evidence suggests that habitat-based analyses of intertumoral heterogeneity can enhance the prediction of neoadjuvant therapy efficacy.⁴⁶ Specifically, prediction models incorporating habitat imaging features demonstrate superior performance compared to traditional radiomics approaches, underscoring the potential of habitat-based characterization to capture the spatial complexity of tumors and refine clinical decision-making.⁴⁷ A key innovation of this study is the

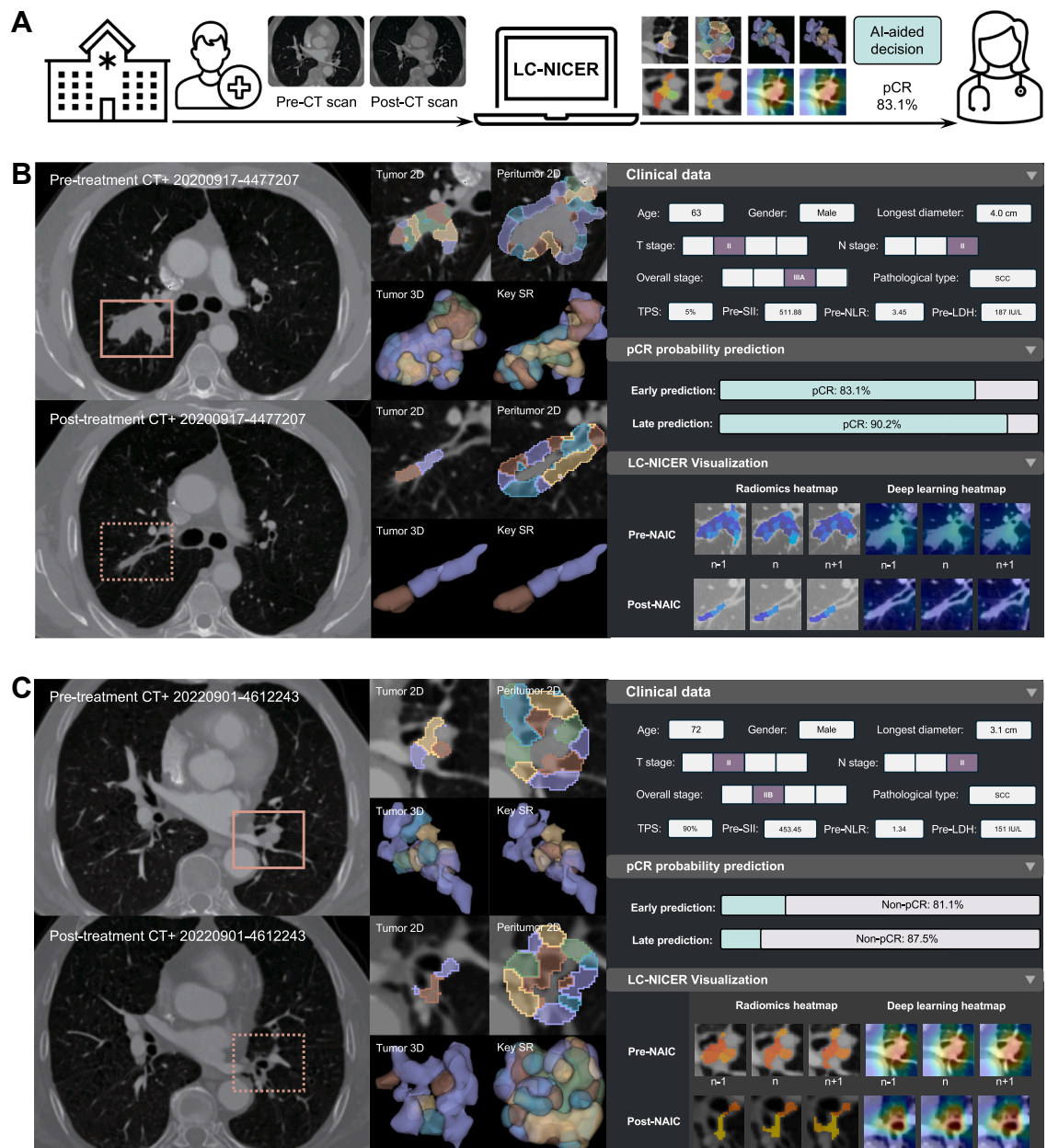


Fig. 7: Clinical decision-support application of LC-NICER. **A.** clinical applicability of the LC-NICER in predicting treatment response. **B.** A 63-year-old male patient with stage IIIA squamous cell carcinoma (tumor diameter: 4.0 cm, TPS: 5%) was classified as non-CCR by RECIST. The LC-NICER_α predicted a pCR probability of 83.1% at the early stage, while LC-NICER_β increased this to 90.2% at the late stage. Post-surgical pathology confirmed pCR. **C.** A 64-year-old male patient with stage IIIB squamous cell carcinoma (tumor diameter: 3.1 cm, TPS: 70%) was classified as CCR by RECIST. The LC-NICER_α and LC-NICER_β predicted non-pCR probabilities of 81.1% and 87.5%, respectively, which were confirmed by post-surgical pathology. The framework provides comprehensive visualizations, including tumor and peritumoral subregions (2D and 3D) and radiomics/deep learning feature heatmaps, highlighting suspicious non-pCR regions to support clinical decision-making. Abbreviations: TPS, Tumor cell Proportion Score; CCR, Clinical Complete Response; pCR, pathological complete response.

development of a novel habitat imaging analysis pipeline, which segments tumors and peritumoral regions into distinct habitats and quantifies their spatial heterogeneity. This approach, combined with radiomics and deep learning features, enables a detailed

characterization of the tumor microenvironment and its changes during treatment.

Despite its promising results, our study has limitations. First, as a retrospective analysis, it is susceptible to selection bias. Compared with similar studies

(sample sizes ranging from $n = 216$ to 509),^{28,48,49} though the largest in this field to date, requires further validation through large-scale prospective clinical trials. Notably, deep learning features underperformed compared to radiomics and habitat imaging, likely due to the lack of task-specific fine-tuning limited by sample size and heterogeneity. Future work will focus on fine-tuning the foundation model to address this. Second, our system currently relies solely on CT imaging features, and the clinical interpretability of deep learning-derived features remains limited. Future studies could enhance predictive performance by integrating multi-omics data (e.g., pathomics, genomics, transcriptomics, and proteomics) to uncover the biological underpinnings of imaging features and further improve the model's accuracy and interpretability. Third, while this study focused exclusively on pCR prediction, future work should incorporate the full spectrum of treatment responses to develop clinically actionable stratification including major, partial, and non-responders. Fourthly, the lack of validation of our model's performance using survival data is another limitation of our research, we will continue to improve follow-up data and further validate the prediction of survival in the LC-NICER model.

In summary, LC-NICER introduces an innovative multi-dimensional framework for predicting NACI response through comprehensive quantification of tumor spatiotemporal heterogeneity, integrating radiomic signatures, deep learning features, and habitat imaging patterns. While immediate widespread adoption faces challenges, this AI-derived longitudinal and multi-dimensional imaging decision system demonstrates significantly enhanced predictive accuracy compared to conventional biomarkers while generating clinically translatable decision-support tools for personalized non-small cell lung cancer neoadjuvant chemoimmunotherapy.

Contributors

Z.S., Z.L., Y.Q., H.Z. ensured the integrity of the research. G.Y. and Z.W. accessed and verified the data. G.Y., Z.W., C.H., G.W., J.G., C.L., Y.L., L.H., L.W., D.R. and A.D. contributed to the conceptualization and design of the study. G.Y., Z.W., C.W., Y.L., X.C. and W.Z. performed data collection and analysis. C.W., Y.L., W.Z. conducted literature review. All authors participated in the writing and revision of the manuscript. All authors approved the final version of the manuscript.

Data sharing statement

The multicenter lung CT datasets analyzed in this study are not publicly available due to patient privacy protections but may be requested for academic research purposes by contacting the corresponding author. The complete analysis pipeline was developed in Python. The proposed LC-NICER software and source code are publicly available at GitHub: <https://github.com/zhenweishi/LC-NICER>.

Declaration of interests

All authors declare no competing interests.

Acknowledgements

G. Y., Z. W., C. H. and G. W. contributed equally to this work as first authors. This work was supported by the Noncommunicable Chronic

Diseases-National Science and Technology Major Project (2024ZD0531100, 2024ZD0531101), the National Natural Science Foundation of China (82472062, 62102103, 82473298, 82202148, 82502479), the National Key Research and Development Program of China (2023YFC2508603), the Natural Science Foundation of Guangdong Province of China (2024A1515011672), Regional Innovation and Development Joint Fund of National Natural Science Foundation of China (U22A20345), Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (2022B1212010011). The authors gratefully acknowledge MediAI Hub (an advanced medical image analysis platform developed and maintained by MediaLab) for its comprehensive technical support, including data curation, image annotation, analysis, and AI model development. The authors also acknowledge the use of PyHiomics toolkit, which enabled the quantification of tumor heterogeneity based on medical imaging data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2025.103551>.

References

- Awad MM, Forde PM, Girard N, et al. Neoadjuvant nivolumab plus ipilimumab versus chemotherapy in resectable lung cancer. *J Clin Oncol*. 2025;43:JCO2402239.
- Lu S, Zhang W, Wu L, et al. Perioperative toripalimab plus chemotherapy for patients with resectable non-small cell lung cancer: the neotorch randomized clinical trial. *JAMA*. 2024;331(3):201–211.
- Cascone T, Awad MM, Spicer JD, et al. Perioperative nivolumab in resectable lung cancer. *N Engl J Med*. 2024;390(19):1756–1769.
- Wakelee H, Liberman M, Kato T, et al. Perioperative pembrolizumab for early-stage non-small-cell lung cancer. *N Engl J Med*. 2023;389(6):491–503.
- Heymach JV, Harpole D, Mitsudomi T, et al. Perioperative durvalumab for resectable non-small-cell lung cancer. *N Engl J Med*. 2023;389(18):1672–1684.
- Provencio M, Nadal E, Insa A, et al. Perioperative chemotherapy and nivolumab in non-small-cell lung cancer (NADIM): 5-year clinical outcomes from a multicentre, single-arm, phase 2 trial. *Lancet Oncol*. 2024;25(11):1453–1464.
- Havel JJ, Chowell D, Chan TA. The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 2019;19(3):133–150.
- Phillips WJ, Camidge DR, Wheatley-Price P. Adjuvant immunotherapy should not be used in patients with a pathologic complete response to neoadjuvant chemoimmunotherapy. *J Thorac Oncol*. 2025;20(1):30–33.
- Smith JJ, Strombom P, Chow OS, et al. Assessment of a watch-and-wait strategy for rectal cancer in patients with a complete response after neoadjuvant therapy. *JAMA Oncol*. 2019;5(4):e185896.
- Rullier E, Rouanet P, Tuech J-J, et al. Organ preservation for rectal cancer (GRECCAR 2): a prospective, randomised, open-label, multicentre, phase 3 trial. *Lancet*. 2017;390(10093):469–479.
- van der Wilk BJ, Eyck BM, Wijnhoven BPL, et al. Neoadjuvant chemoradiotherapy followed by active surveillance versus standard surgery for oesophageal cancer (SANO trial): a multicentre, stepped-wedge, cluster-randomised, non-inferiority, phase 3 trial. *Lancet Oncol*. 2025;26(4):425–436.
- Herbst RS, Baas P, Kim D-W, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016;387(10027):1540–1550.
- Mino-Kenudson M, Schalper K, Cooper W, et al. Predictive biomarkers for immunotherapy in lung cancer: perspective from the international association for the study of lung cancer pathology committee. *J Thorac Oncol*. 2022;17(12):1335–1354.
- Hwang DM, Alabaqer T, Santiago RC, et al. Prevalence and heterogeneity of PD-L1 expression by 22C3 assay in routine population-based and reflexive clinical testing in lung cancer. *J Thorac Oncol*. 2021;16(9):1490–1500.
- Liu Z, Yang Z, Wu J, et al. A single-cell atlas reveals immune heterogeneity in anti-PD-1-treated non-small cell lung cancer. *Cell*. 2025;188(11):3081–3096.e19.

- 16 Provencio M, Nadal E, Insa A, et al. Neoadjuvant chemotherapy and nivolumab in resectable non-small-cell lung cancer (NADIM): an open-label, multicentre, single-arm, phase 2 trial. *Lancet Oncol*. 2020;21(11):1413–1422.
- 17 Cheng Y, Jiang Z-P, Chen X-B, Lu K-Y, Liu Z-Y, Shao D. Prognostic value of 18F-FDG PET/CT metabolic parameters in resectable non-small cell lung cancer treated with neoadjuvant immunotherapy plus chemotherapy. *Clin Nucl Med*. 2025;50(6):e344–e351.
- 18 Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. *Nat Rev Cancer*. 2016;16(5):275–287.
- 19 Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
- 20 Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. *N Engl J Med*. 2023;388(21):1981–1990.
- 21 Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773–1784.
- 22 Perez-Lopez R, Ghaffari Laleh N, Mahmood F, Kather JN. A guide to artificial intelligence for cancer researchers. *Nat Rev Cancer*. 2024;24(6):427–441.
- 23 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.
- 24 Ye G, Wu G, Qi Y, et al. Non-invasive multimodal CT deep learning biomarker to predict pathological complete response of non-small cell lung cancer following neoadjuvant immunochemotherapy: a multicenter study. *J Immunother Cancer*. 2024;12(9):e009348.
- 25 Huang Y, Zhu T, Zhang X, et al. Longitudinal MRI-based fusion novel model predicts pathological complete response in breast cancer treated with neoadjuvant chemotherapy: a multicenter, retrospective study. *eClinicalMedicine*. 2023;58:101899.
- 26 Li Z, Gao J, Zhou H, et al. Multiregional dynamic contrast-enhanced MRI-based integrated system for predicting pathological complete response of axillary lymph node to neoadjuvant chemotherapy in breast cancer: Multicentre study. *eBioMedicine*. 2024;107:105311.
- 27 Huang Y-H, Shi Z-Y, Zhu T, et al. Longitudinal MRI-driven multi-modality approach for predicting pathological complete response and B cell infiltration in breast cancer. *Adv Sci (Weinh)*. 2025;12(12):e2413702.
- 28 Zheng J, Yan Z, Wang R, et al. NeoPred: dual-phase CT AI forecasts pathologic response to neoadjuvant chemo-immunotherapy in NSCLC. *J Immunother Cancer*. 2025;13(5):e011773.
- 29 Han X, Wang M, Zheng Y, et al. Delta-radiomics features for predicting the major pathological response to neoadjuvant chemoimmunotherapy in non-small cell lung cancer. *Eur Radiol*. 2024;34(4):2716–2726.
- 30 Liu Y, Wang Y, Wang Y, et al. Early prediction of treatment response to neoadjuvant chemotherapy based on longitudinal ultrasound images of HER2-positive breast cancer patients by Siamese multi-task network: a multicentre, retrospective cohort study. *eClinicalMedicine*. 2022;52:101562.
- 31 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
- 32 Fong KM, Rosenthal A, Giroux DJ, et al. The international association for the study of lung cancer staging project for lung cancer: proposals for the revision of the M descriptors in the forthcoming ninth edition of the TNM classification for lung cancer. *J Thorac Oncol*. 2024;19(5):786–802.
- 33 Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228–247.
- 34 Mountzios G, Remon J, Hendriks LEL, et al. Immune-checkpoint inhibition for resectable non-small-cell lung cancer - opportunities and challenges. *Nat Rev Clin Oncol*. 2023;20(10):664–677.
- 35 Spicer JD, Cascone T, Wynes MW, et al. Neoadjuvant and adjuvant treatments for early stage resectable NSCLC: consensus recommendations from the international association for the study of lung cancer. *J Thorac Oncol*. 2024;19(10):1373–1414.
- 36 McLaughlin J, Han G, Schalper KA, et al. Quantitative assessment of the heterogeneity of PD-L1 expression in non-small-cell lung cancer. *JAMA Oncol*. 2016;2(1):46–54.
- 37 Borcoman E, Kanjanapan Y, Champiat S, et al. Novel patterns of response under immunotherapy. *Ann Oncol*. 2019;30(3):385–396.
- 38 Shi L, Li C, Bai Y, et al. CT radiomics to predict pathologic complete response after neoadjuvant immunotherapy plus chemoradiotherapy in locally advanced esophageal squamous cell carcinoma. *Eur Radiol*. 2025;35(3):1594–1604.
- 39 Saad MB, Hong L, Aminu M, et al. Predicting benefit from immune checkpoint inhibitors in patients with non-small-cell lung cancer by CT-based ensemble deep learning: a retrospective study. *Lancet Digit Health*. 2023;5(7):e404–e420.
- 40 Lin P, Xie W, Li Y, et al. Intratumoral and peritumoral radiomics of MRIs predicts pathologic complete response to neoadjuvant chemoimmunotherapy in patients with head and neck squamous cell carcinoma. *J Immunother Cancer*. 2024;12(11):e009616.
- 41 Lin Q, Wu HJ, Song QS, Tang YK. CT-based radiomics in predicting pathological response in non-small cell lung cancer patients receiving neoadjuvant immunotherapy. *Front Oncol*. 2022;12:937277.
- 42 Qu W, Chen C, Cai C, et al. Non-invasive prediction for pathologic complete response to neoadjuvant chemoimmunotherapy in lung cancer using CT-based deep learning: a multicenter study. *Front Immunol*. 2024;15:1327779.
- 43 Ye G, Wu G, Zhang C, et al. CT-based quantification of intratumoral heterogeneity for predicting pathologic complete response to neoadjuvant immunochemotherapy in non-small cell lung cancer. *Front Immunol*. 2024;15:1414954.
- 44 Mazzaschi G, Marrochio C, Moron Dalla Tor L, et al. Longitudinal blood immune-inflammatory and radiomic profiling to decode different patterns of acquired resistance to immunotherapy in patients with NSCLC. *Clin Cancer Res*. 2025;31(8):1533–1545.
- 45 Jin C, Yu H, Ke J, et al. Predicting treatment response from longitudinal images using multi-task deep learning. *Nat Commun*. 2021;12(1):1851.
- 46 Shi Z, Huang X, Cheng Z, et al. MRI-based quantification of intratumoral heterogeneity for predicting treatment response to neoadjuvant chemotherapy in breast cancer. *Radiology*. 2023;308(1):e222830.
- 47 Sujit SJ, Aminu M, Karpins TV, et al. Enhancing NSCLC recurrence prediction with PET/CT habitat imaging, ctDNA, and integrative radiogenomics-blood insights. *Nat Commun*. 2024;15(1):3152.
- 48 Geng Z, Li K, Mei P, et al. Multichannel deep learning prediction of major pathological response after neoadjuvant immunochemotherapy in lung cancer: a multicenter diagnostic study. *Int J Surg*. 2025. <https://doi.org/10.1097/JS9.0000000000002821>.
- 49 Fan S, Xie J, Zheng S, et al. Non-invasive CT based multiregional radiomics for predicting pathologic complete response to preoperative neoadjuvant chemoimmunotherapy in non-small cell lung cancer. *Eur J Radiol*. 2025;189:112171.