Prognostic and predictive values of a multimodal nomogram incorporating tumor and peritumor morphology with immune status in resectable lung adenocarcinoma

Huan Lin [iD],[1] Junjie Hua [iD],[2] Yumeng Wang,[3] Mingwei Chen,[3] Yanting Liang,[1] LiXu Yan,[4] Wei Zhao,[5] Shiwei Luo,[5] Deqing Hong,[6] Xin Chen,[7] Xipeng Pan,[3] Jun Liu,[5] Zaiyi Liu[1,8]

Correspondence to
Professor Zaiyi Liu;
liuzaiyi@gdph.org.cn

ABSTRACT

Background Current prognostic and predictive biomarkers for lung adenocarcinoma (LUAD) predominantly rely on unimodal approaches, limiting their characterization ability. There is an urgent need for a comprehensive and accurate biomarker to guide individualized adjuvant therapy decisions.

Methods In this retrospective study, data from patients with resectable LUAD (stage I–III) were collected from two hospitals and a publicly available dataset, forming a training dataset (n=223), a validation dataset (n=95), a testing dataset (n=449), and the non-small cell lung cancer (NSCLC) Radiogenomics dataset (n=59). Tumor and peritumor scores were constructed from preoperative CT radiomics features (shape/intensity/texture). An immune score was derived from the density of tumor-infiltrating lymphocytes (TILs) within the cancer epithelium and stroma on hematoxylin and eosin-stained whole-slide images. A clinical score was constructed based on clinicopathological risk factors. A Cox regression model was employed to integrate these scores, thereby constructing a multimodal nomogram to predict disease-free survival (DFS). The adjuvant chemotherapy benefit rate was subsequently calculated based on this nomogram.

Results The multimodal nomogram outperformed each of the unimodal scores in predicting DFS, with a C-index of 0.769 (vs 0.634–0.731) in the training dataset, 0.730 (vs 0.548–0.713) in the validation dataset, and 0.751 (vs 0.660–0.692) in the testing dataset. It was independently associated with DFS after adjusting for other clinicopathological risk factors (training dataset: HR=3.02, p<0.001; validation dataset: HR=2.33, p<0.001; testing dataset: HR=2.03, p=0.001). The adjuvant chemotherapy benefit rate effectively distinguished between patients benefiting from adjuvant chemotherapy and those from observation alone (interaction p<0.001). Furthermore, the high-/low-risk groups defined by the multimodal nomogram provided refined stratification of candidates for adjuvant chemotherapy identified by current guidelines (p<0.001). Gene set enrichment analyses using the NSCLC Radiogenomics dataset revealed associations between tumor/peritumor scores and pathways involved in epithelial-mesenchymal transition, angiogenesis, IL6-JAK-STAT3 signaling, and reactive oxidative species.

Conclusion The multimodal nomogram, which incorporates tumor and peritumor morphology with anti-tumor immune response, provides superior prognostic accuracy compared with unimodal scores. Its defined adjuvant chemotherapy benefit rates can inform individualized adjuvant therapy decisions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Accurate prognostic stratification and personalized adjuvant therapy decisions in resectable lung adenocarcinoma (LUAD) remain challenging. Current biomarkers based on clinical, radiological, and histopathological data offer limited prognostic and predictive accuracy when used in isolation. CT imaging provides information on tumor and peritumor morphology, while the quantification of tumor-infiltrating lymphocytes (TILs) from histopathological images reflects the antitumor immune response. These routinely available data hold promises for complementing each other to enable a more comprehensive characterization of LUAD.

BACKGROUND

Lung cancer represents the malignancy with the highest incidence and mortality worldwide.[1] Lung adenocarcinoma (LUAD) is the most common subtype,[2] with 25–30% of cases being resectable. Nonetheless, surgical resection alone is not always curative. Some of these patients eventually experience recurrence, metastasis, or death.[3] The tumor-node-metastasis (TNM) staging system[4] is

WHAT THIS STUDY ADDS

⇒ This study introduces a novel multimodal nomogram that integrates tumor and peritumor scores from preoperative CT radiomics, an immune score derived from the density of TILs within the cancer epithelium and stroma on hematoxylin and eosin-stained whole-slide images, and a clinical score based on clinicopathological risk factors. The nomogram demonstrated superior accuracy in predicting disease-free survival across multicenter datasets compared with individual unimodal scores. Furthermore, its defined adjuvant chemotherapy benefit rate effectively distinguished patients who would benefit from adjuvant chemotherapy vs observation alone, providing refined stratification and supporting personalized adjuvant therapy decisions.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ This study offers a practical framework for multimodal integration using routine clinical data. By elucidating the histopathological and biological basis of radiomics-based tumor and peritumor scores and investigating the association between immune score and immune phenotypes, the nomogram's interpretability is enhanced. This deeper understanding of how multimodal data provide complementary insights leads to improved prognostic and predictive accuracy. The multimodal nomogram and its derived adjuvant chemotherapy benefit rate can refine postoperative adjuvant therapy decisions for resectable LUAD, complementing current guidelines and further individualizing treatment to minimize both overtreatment and undertreatment.

the standard for risk stratification and treatment guidance for LUAD. The National Comprehensive Cancer Network (NCCN)[5] and the European Society for Medical Oncology (ESMO)[6] guidelines also list high-risk features associated with recurrence, such as tumors of ≥4mm, poor differentiation, vascular invasion, wedge resection, visceral pleural involvement, and unknown lymph node status (Nx). However, existing risk stratification and treatment guidelines are often underperforming, as patients with the same TNM stage and high-risk features do not always exhibit identical outcomes.[7] Furthermore, as a common systemic therapy, adjuvant chemotherapy is recommended for stage IB–IIA patients with high-risk features and stage IIB–III patients. However, it improves survival by only 5–10%,[8] resulting in potential overtreatment. It is, therefore, necessary to develop more accurate prognostic models to support individualized postoperative treatment decisions.

Patients with resectable LUAD are required to undergo contrast-enhanced CT to assess the tumor and surrounding tissues for surgical planning. The morphological characteristics in these images provide valuable prognostic information.[9 10] Compared with conventional manual interpretation, radiomics approaches analyze imaging quantitatively, enabling more comprehensive identification of prognostically relevant morphological patterns. Vaidya et al[11] developed a radiomic risk score and associated nomogram based on tumor/peritumoral features and high-risk features to predict disease-free survival (DFS) in stage I–II non-small cell lung cancer

(NSCLC), achieving C-indexes of 0.72–0.74 across multiple cohorts. To address the interpretability challenges in radiomics models, some researchers explored the underlying biological mechanisms.[11 12] However, CT imaging cannot directly provide biological information at the microscopic scale due to its imaging principle. Therefore, further incorporating the biological function of LUAD is necessary to offer comprehensive information for accurate prognostication.

The tumor microenvironment influences the biological processes involved in the development, progression, and prognosis of LUAD. As a significant component of the microenvironment, tumor-infiltrating lymphocytes (TILs) contribute to the antitumor immune responses.[13] A high density of TILs is confirmed as a favorable prognostic factor[14] and also a predictor of responders to immune checkpoint inhibitors.[15] Over the past decade, researchers have proposed implementing a TIL-based immunoscore into clinical practice as a supplement to the TNM stage.[16] Immunoscore is often constructed by incorporating the density of TILs in both cancer epithelium and stroma[16] or in the invasive margin and tumor center,[17 18] due to the distinct prognostic implications of these compartments.[18 19] Hematoxylin and eosin (H&E)-stained histopathological sections, sampled from surgically resected tissue, are the most efficient and convenient method for TILs quantification. Recent studies[15 20] have digitized the sections into whole-slide images (WSIs) and identified TILs in cancer epithelium and stroma using artificial intelligence algorithms, significantly improving efficiency and reproducibility compared with manual evaluation.[21] Pan et al[22] developed a WELL score based on automated quantification of TIL density using deep learning methods and validated its prognostic value across multicenter LUAD cohorts. Although the performance of the WELL score is robust (C-index 0.53–0.62), its accuracy still needs improvement.

In recent years, multimodal integration analysis has been gradually applied in various malignancies, including pathology-genomics,[23 24] radiology-pathology,[25] and radiology-pathology-genomics.[26 27] The findings indicate that compared with unimodal approaches, multimodal integration demonstrates superior and robust performance in risk stratification and treatment responder identification.[28 29] A recent study by Chen et al[24] demonstrated that histology-genomics integration improved the risk stratification of LUAD. Compared with genomics data, radiological and pathological images are routine clinical data, possessing advantages in cost-effectiveness and availability, and offering complementary information on macroscopic and microscopic scales. However, currently, no attempt has been made to incorporate radiological and histopathological images for LUAD prognostic stratification.

This study aims to develop a multimodal nomogram by incorporating tumor/peritumor scores derived from CT images, an immune score computed from pathological WSIs, and a clinical score based on clinical information,
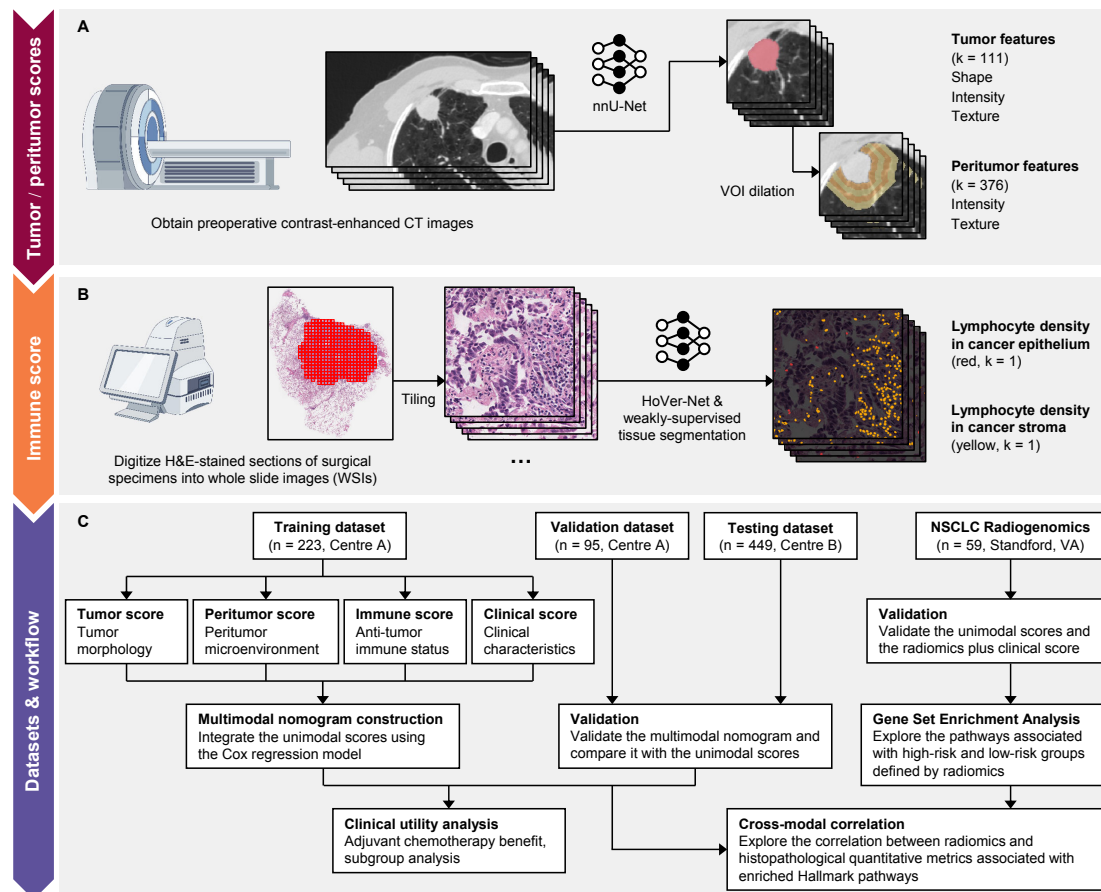
Figure 1 Study design and workflow. (A) Contrast-enhanced chest CT images are routinely obtained from patients with lung adenocarcinoma (LUAD). Tumor volumes of interest (VOIs) are segmented with the assistance of nnU-Net, while peritumor VOIs are generated using a dilation algorithm based on tumor VOIs. Radiomics features are extracted separately from the tumor and peritumor VOIs, and tumor and peritumor scores are constructed using the selected representative features. (B) H&E-stained sections of surgical specimens, which are routine data for LUAD patients, are digitized into whole-slide images. Lymphocytes within the cancer epithelium and stroma compartments are identified using deep-learning models. An immune score is then derived from the lymphocyte density within each compartment. (C) The multimodal nomogram is constructed by integrating tumor, peritumor, immune, and clinical scores using a Cox regression model. The prognostic and predictive value of the nomogram is validated using multicenter datasets. The histopathological and biological basis of the radiomics-based tumor and peritumor scores is further explored through gene set enrichment analysis and cross-modal correlation analysis.

characterizing the morphology and antitumor immune response of LUAD. The prognostic and predictive validity of the multimodal nomogram was validated using multi-center and publicly available NSCLC Radiogenomics datasets. We further explored the histopathological characteristics and gene pathways related to the tumor/peritumor score to elucidate its underlying mechanisms, thereby enhancing our understanding of how CT images and the immune score complement each other to improve prognostic capabilities.

METHODS
Patient enrollment and clinical information collection
This study was conducted at two tertiary hospitals in China, referred to here as Centers A and B. The ethics committees of both hospitals approved this retrospective study (approval numbers: KY-Q-2022-338-01 and 2021557), waiving the requirement for informed consent.

The study workflow is illustrated in figure 1. The study subjects were patients with invasive LUAD who had undergone curative surgery and had contrast-enhanced chest CT images (slice thickness≤3 mm) within 90 days before surgery, with H&E-stained histopathological sections of surgical specimens available. A total of 409 patients from Center A (November 2007 to December 2014) and 639 patients from Center B (August 2013 to December 2017) were included consecutively. Patients were excluded if they had undergone incomplete resection (R1/R2), experienced recurrence or stage IV disease, had adenocarcinoma in situ, had poor-quality CT images or WSIs if the lesion analyzed on CT was not the same as the lesion displayed on the WSI, had a history of malignancies, received neoadjuvant therapy, or died within 30 days after surgery. The inclusion and exclusion flowchart is presented in online supplemental figure S1.

The study endpoint was DFS, defined as the time elapsed between the date of surgery and the date of the first recurrence or death. Clinical data were collected from medical records, and all pathological TNM (pTNM) stages were manually verified to guarantee consistency with the eighth edition of the American Joint Committee on Cancer staging system. Patients lacking essential information, such as pTNM stage or treatment data, were excluded, and no imputation was performed for missing data.

Furthermore, the radiomics-based tumor and peritumor scores were validated using invasive LUAD cases from the publicly available NSCLC Radiogenomics dataset (https://wiki.cancerimagingarchive.net/display/Public/NSCLC+Radiogenomics). The prognostic performance of the radiomics plus clinical score model was evaluated, and the biological mechanisms underlying the radiomics-based score were further analyzed using RNA sequencing data from this dataset. Inclusion criteria: invasive LUAD patients who had contrast-enhanced chest CT images (slice thickness ≤3 mm) before surgery. Exclusion criteria: stage IV disease, adenocarcinoma in situ, death within 30 days after surgery, and lack of essential information (such as pTNM stage).

Chest CT image acquisition and annotation

The contrast-enhanced chest CT images used in this study were obtained from routine clinical procedures for LUAD patients (figure 1A). The CT imaging protocols employed at Centers A and B are detailed in online supplemental table S1. To guarantee the quality, all CT images were manually reviewed by a radiologist with 9 years of experience (HL) in chest imaging.

A semi-automated segmentation method was employed to enhance the efficiency and accuracy of the annotation of volumes of interest (VOIs) for tumor and peritumor regions. (1) Pre-segmentation of tumor VOIs: the initial segmentation of tumor VOIs was conducted using a pre-trained nnU-Net[30] model. This model was trained on CT images (with VOIs annotated by radiologists) of over 1000 LUAD cases from Center A. (2) Manual inspection and correction of tumor VOIs: a radiologist (HL) conducted a manual inspection and correction of the pre-segmented tumor VOIs when necessary. (3) Pre-segmentation of peritumor VOIs: based on the tumor VOIs, a dilation algorithm was applied to generate peritumor VOIs in the lung parenchyma at distances of 0–5 mm, 5–10 mm, 10–15 mm, and 15–20 mm surrounding the tumor. (4) Manual inspection and correction of peritumor VOIs: a radiologist (HL) conducted a manual inspection and correction of the pre-segmented peritumor VOIs, ensuring the exclusion of chest wall and mediastinal structures.

To evaluate the robustness of radiomics features across VOIs delineated by different annotators, 130 cases from the training dataset were randomly selected, and the VOIs were delineated manually by another radiologist with 3 years of experience in chest imaging (YL). VOI delineation and corrections were performed using ITK-SNAP 3.8.0.

Extraction of radiomics features and construction of the tumor/peritumor scores

A total of 487 radiomics features were extracted from the tumor and peritumor VOIs using PyRadiomics 3.0.1. These features are detailed in online supplemental table S2, which characterized the shape, intensity, and texture of the tumor regions, as well as the intensity and texture of the peritumor regions.

All radiomic features were z-score normalized, and the interclass correlation coefficient (ICC) was calculated to identify those demonstrating robustness (ICC >0.75) across different annotators' delineations of VOIs. Subsequently, if two features exhibited collinearity (Pearson correlation coefficient >0.9), the feature with clearer medical significance and greater stability was retained for analysis. LASSO Cox regression was employed for feature selection, with the penalty parameter determined through 10-fold cross-validation using the lambda.min rule. A total of 1000 times of 10-fold cross-validation LASSO Cox regression were performed, with each time involving all features and all samples. In each 10-fold cross-validation process, the samples were divided into 10 equal-sized subsets, with each subset used as a validation set while the remaining subsets served as training sets. Features with nonzero coefficients in each LASSO Cox regression were retained, and the frequency of each feature being selected across the 1000 times was calculated. The selection frequency of each feature was recorded to assess its importance. Tumor features selected over 850 times and peritumor features selected over 999 times were considered candidate features for constructing the tumor score and peritumor score. Finally, a stepwise Cox regression analysis based on the Akaike Information Criterion (AIC) was performed to determine the features that constituted the tumor score and peritumor score. The linear predictor of the Cox regression model served as the score value.

Histopathological WSI acquisition and TIL identification

The H&E-stained paraffin sections of surgical specimens were obtained as routine data from resectable LUAD patients. A pathologist with 15 years of experience in thoracic pathology (LY) selected a section that best represented the entire tumor for each patient, without knowledge of the patient's outcomes. The selected section was ensured to be complete and free of large necrotic areas. The sections were digitized into WSIs using whole-slide scanners (figure 1B), with a resolution of 0.252 μm/pixel in Center A (Leica Aperio AT2, USA) and 0.264 μm/pixel in Center B (Leica Aperio GT 450, USA). All WSIs were manually inspected to ensure quality, with areas exhibiting over-light or over-dark staining, tissue cracks or folds, insufficient tissue, air bubbles, blurring, or streaking being excluded.

The quantification of TIL density in the cancer epithelium and stroma required the segmentation of cancer epithelium and stroma tissue, as well as the identification of TILs in each compartment. To this end, a series of deep-learning models were employed for the automated segmentation and identification of tumor regions,[31] cancer epithelium and stroma tissues,[32] and lymphocyte nuclei,[31 33] as detailed in online supplemental A1.

Construction of the immune score and calculation of histopathological quantitative metrics

In consideration of the discrepancy in prognostic relevance between the density of TILs in the cancer epithelium and stroma, we constructed the immune score by combining TILs densities (cells/mm²) in these two compartments using a Cox regression model, referencing Pan et al,[22] with the linear predictor of the model serving as the immune score value.

Furthermore, the following histopathological metrics were quantified based on the segmentation results: (1) stroma fraction[34]=area of cancer stroma / area of cancer epithelium and stroma. (2) Necrosis ratio[35]=area of necrotic region / area of tumor region (applicable only to cases with necrotic regions). (3) Cellularity[36]=density of cancer cells within the tumor region × maximum diameter of tumor region. (4) Immune phenotypes[15]: based on the distribution of TILs density within the cancer epithelium, stroma, and combined regions of both, immune phenotypes were classified as inflamed, immune-excluded, or immune-desert,[37] according to the training dataset. Those with TIL density in the lowest third of the combined regions were classified as immune-desert. Those above this threshold were classified as either inflamed or immune-excluded, with further classification based on the median of the ratio of TIL density in the cancer epithelium to that in the stroma.

Gene set enrichment analysis (GSEA) in the NSCLC Radiogenomics dataset

In the NSCLC Radiogenomics dataset, patients with both tumor and peritumor scores in the lower 50% were defined as the low-risk group, while those with both tumor and peritumor scores in the upper 50% were defined as the high-risk group. The difference in DFS between the low-risk and high-risk groups was assessed.

Subsequently, GSEA was performed using RNA sequencing data from these patients (GSE103584, Illumina HiSeq 2500) to elucidate the biological mechanisms associated with these radiomics-based unimodal scores. Differentially expressed genes between the high-risk and low-risk groups were analyzed using the GEO2R tool from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/geo/geo2r/). The resulting data were then ranked based on the logarithm of fold change in differential gene expression levels. GSEA was conducted using the R package clusterProfiler.[38] The hallmark gene sets from the Molecular Signatures Database[39] were referenced, and the false discovery rate (FDR) was applied for multiple comparison correction, with gene sets demonstrating an FDR of <0.05 considered significantly enriched.

Statistical analysis

Before constructing the clinical score, we first confirmed that each variable met the proportional hazards assumption using log-log plots. Clinical risk factors were selected in the training dataset using univariable Cox regression analysis with a p<0.20 to avoid missing important factors. Stepwise Cox regression analysis based on AIC with a p<0.05 was subsequently employed to identify the variables for constructing the clinical score. The linear predictor of the model was used as the value of the clinical score.

Based on the training dataset, a Cox regression model was constructed to integrate the unimodal scores into a multimodal nomogram for predicting DFS. The optimal cut-off point was determined using X-tile 3.6.1 (Yale University, New Haven, CT, USA),[40] which divided patients into low-risk and high-risk groups. X-tile employs an enumeration method, in which various values of the continuous variable are tested as potential cut-off points. A log-rank test is then conducted to compare survival curves between the resulting groups, with the cut-off point corresponding to the smallest p value being selected as the optimal value. The Kaplan–Meier method was employed to evaluate the difference in DFS between the two groups in the validation and testing datasets. A multivariable Cox regression analysis was employed to assess whether the association between the multimodal nomogram and DFS was independent of other clinicopathological risk factors. The performance of the unimodal scores and the multimodal nomogram was evaluated in terms of calibration and discrimination. Calibration was assessed using AIC values, while discrimination was evaluated using time-dependent area under the receiver operating characteristics curve (AUC) and Harrell's concordance index (C-index). The enhancement of the multimodal nomogram compared with the unimodal scores was quantified using net reclassification improvement (NRI) and integrated discrimination improvement (IDI).

To gain insight into the histopathological basis of the radiomics-based tumor and peritumor scores, a Spearman correlation analysis was performed to analyze the association between the scores and histopathological quantitative metrics (stroma fraction, necrosis ratio, and cancer cellularity). A Kruskal–Wallis H test was conducted to compare the differences in immune score across the different immune phenotypes, namely, inflamed, immune-excluded, and immune-desert.

The statistical analyses were conducted using R 4.0.3 (R Foundation, Vienna, Austria) and SPSS 27.0 (IBM, Armonk, NY, USA). Unless otherwise stated, a two-sided p<0.05 was considered statistically significant. The sample size for this retrospective study was predetermined, with 89 DFS events in the training dataset. We ensured that the number of candidate predictor parameters did not

exceed nine, following the rule of thumb for sufficient sample size.

RESULTS

Baseline clinicopathological characteristics of LUAD patients

Based on the inclusion and exclusion criteria, Center A ultimately included 318 patients, who were then divided into a training dataset (n=223) and a validation dataset (n=95) in a 7:3 ratio according to the surgery date. Center B ultimately included 449 patients, forming the testing dataset. Additionally, 59 patients were included in the NSCLC Radiogenomics dataset. Table 1 presents the baseline clinicopathological characteristics of LUAD patients across the four datasets.

Prognostic value of each unimodal score

Tumor score

Among the 487 radiomic features, 479 demonstrated robustness across VOIs delineated by different annotators. From the tumor features, 38 independent features were retained. Through 1000 times of LASSO Cox regression, three features were selected as candidates, followed by stepwise Cox regression based on AIC to optimize the feature set. Ultimately, two texture features (RaTumor_original_glrlm_LongRunHighGrayLevelEmphasis and RaTumor_original_gldm_DependenceVariance) were selected to construct the tumor score. Patients with higher tumor scores had worse DFS in the training dataset (HR=2.72, 95% CI 1.93 to 3.83, p<0.001), validation dataset (HR=4.13, 95% CI 2.46 to 6.94, p<0.001), and testing dataset (HR=3.21, 95% CI 2.05 to 5.04, p<0.001) (online supplemental table S3). The prognostic performance was assessed by C-index, with values of 0.705 (95% CI 0.647 to 0.762) in the training dataset, 0.713 (95% CI 0.641 to 0.785) in the validation dataset, and 0.661 (95% CI 0.601 to 0.721) in the testing dataset.

Peritumor score

Similar strategies were applied for selecting peritumor features, resulting in 70 independent features. After 1000 times of LASSO Cox regression, four features were identified as candidates, and stepwise Cox regression based on AIC was used to refine the feature set. Three intensity/texture features (RaPeritumoral_05 mm_original_gldm_LargeDependenceLowGrayLevelEmphasis, RaPeritumoral_20 mm_original_gldm_GrayLevelNonUniformity, RaPeritumoral_05 mm_original_firstorder_Skewness) were selected for constructing the peritumor score. Patients with higher peritumor scores had worse DFS in the training dataset (HR=2.72, 95% CI 2.03 to 3.64, p<0.001), validation dataset (HR=2.15, 95% CI 1.65 to 2.81, p<0.001), and testing dataset (HR=2.47, 95% CI 1.88 to 3.24, p<0.001) (online supplemental table S3). The C-index for the peritumor score was 0.722 (95% CI 0.667 to 0.777) in the training dataset, 0.713 (95% CI 0.640 to 0.786) in the validation dataset, and 0.692 (95% CI 0.629 to 0.754) in the testing dataset.

Immune score

The immune score was constructed by integrating TIL densities within both the cancer epithelium and stroma compartments. Patients with higher immune scores had worse DFS in the training dataset (HR=2.72, 95% CI 1.60 to 4.60, p<0.001) and testing dataset (HR=3.09, 95% CI 1.64 to 5.83, p<0.001); however, this trend was not significant in the validation dataset (HR=1.32, 95% CI 0.70 to 2.47, p=0.391) (online supplemental table S3). As a result, the predictive performance of the immune score varied across the datasets, with C-index values of 0.634 (95% CI 0.574 to 0.693) in the training dataset, 0.548 (95% CI 0.466 to 0.630) in the validation dataset, and 0.660 (95% CI 0.601 to 0.719) in the testing dataset.

Clinical score

Through stepwise Cox regression based on AIC, variables including age (≥65 vs <65), pT stage, pN stage, and vascular invasion were selected to construct the clinical score (online supplemental table S4). Higher clinical scores were associated with worse DFS in the training dataset (HR=2.72, 95% CI 2.07 to 3.57, p<0.001), validation set (HR=2.06, 95% CI 1.45 to 2.92, p<0.001), and testing dataset (HR=2.31, 95% CI 1.75 to 3.04, p<0.001) (online supplemental table S3). The C-index for the clinical score was 0.731 (95% CI 0.679 to 0.782) in the training dataset, 0.655 (95% CI 0.574 to 0.735) in the validation dataset, and 0.684 (95% CI 0.627 to 0.740) in the testing dataset.

The formulas of tumor, peritumor, immune, and clinical scores are presented in online supplemental A2. The variable coding, partial regression coefficient, and estimated 3 year baseline cumulative hazard of unimodal scores are presented in online supplemental table S5.

Prognostic value of the multimodal nomogram

Based on the training dataset, a Cox regression model was employed to integrate the linear predictors from the tumor score, peritumor score, immune score, and clinical score, thereby constructing a multimodal nomogram for predicting DFS in resectable LUAD (online supplemental figure S2).

A higher multimodal nomogram score was significantly associated with worse DFS across all datasets (training dataset: HR=2.72, 95% CI 2.14 to 3.44, p<0.001; validation dataset: HR=2.21, 95% CI 1.67 to 2.92, p<0.001; testing dataset: HR=2.62, 95% CI 2.07 to 3.33, p<0.001, table 2). The multimodal nomogram demonstrated superior prognostic capability compared with any unimodal score (figure 2A), with a C-index of 0.769 (95% CI 0.719 to 0.819) in the training dataset, 0.730 (95% CI 0.659 to 0.801) in the validation dataset, and 0.751 (95% CI 0.699 to 0.803) in the testing dataset, as evidenced by positive NRIs and IDIs (mostly p<0.05, online supplemental table S3). The time-dependent AUCs for the multimodal nomogram in predicting 2, 3, 4, and 5 year DFS

Table 1 Baseline clinicopathological characteristics of lung adenocarcinoma patients in the training, validation, testing, and non-small cell lung cancer (NSCLC) Radiogenomics datasets

Characteristics	Training dataset n=223	Validation dataset n=95	Testing dataset n=449	NSCLC Radiogenomics n=59	Total n=826
Age (year, mean±SD)	61.3±10.3	62.0±12.3	58.2±8.7	69.3±9.7	60.2±10.1
Age group (%)					
<65	133 (59.6)	57 (60.0)	341 (75.9)	15 (25.4)	546 (66.1)
≥65	90 (40.4)	38 (40.0)	108 (24.1)	44 (74.6)	280 (33.9)
Sex (%)					
Male	111 (49.8)	55 (57.9)	233 (51.9)	42 (71.2)	441 (53.4)
Female	112 (50.2)	40 (42.1)	216 (48.1)	17 (28.8)	385 (46.6)
Smoking status (%)					
Never	168 (75.3)	69 (72.6)	311 (69.3)	11 (18.6)	559 (67.7)
Current or former	55 (24.7)	26 (27.4)	138 (30.7)	48 (81.4)	267 (32.3)
pT stage (%)					
pT1	102 (45.7)	43 (45.3)	156 (34.7)	28 (47.5)	329 (39.8)
pT2	104 (46.6)	43 (45.3)	258 (57.5)	22 (37.3)	427 (51.7)
pT3	15 (6.7)	8 (8.4)	34 (7.6)	8 (13.6)	65 (7.9)
pT4	2 (0.9)	1 (1.1)	1 (0.2)	1 (1.7)	5 (0.6)
pN stage (%)					
pN0	175 (78.5)	64 (67.4)	330 (73.5)	44 (74.6)	613 (74.2)
pN1	10 (4.5)	9 (9.5)	35 (7.8)	6 (10.2)	60 (7.3)
pN2	30 (13.5)	15 (15.8)	72 (16.0)	9 (15.3)	126 (15.3)
pNx	8 (3.6)	7 (7.4)	12 (2.7)	0 (0)	27 (3.3)
pTNM stage* (%)					
Stage I	170 (76.2)	62 (65.3)	299 (66.6)	37 (62.7)	568 (68.8)
Stage II	21 (9.4)	16 (16.8)	82 (16.0)	12 (20.3)	121 (14.6)
Stage III	32 (14.3)	17 (17.9)	78 (17.4)	10 (16.9)	137 (16.6)
Lesion site (%)					
Upper/middle lobe	157 (70.4)	62 (65.3)	296 (65.9)	43 (72.9)	558 (67.6)
Lower lobe	66 (29.6)	33 (34.7)	153 (34.1)	16 (27.1)	268 (32.4)
Differentiation (%)					
Well (G1)/moderate (G2)	178 (79.8)	67 (70.5)	304 (67.7)	45 (76.3)	594 (71.9)
Poor (G3)	41 (18.4)	27 (28.4)	143 (31.8)	14 (23.7)	225 (27.2)
Unknown	4 (1.8)	1 (1.1)	2 (0.4)	0 (0)	7 (0.8)
Visceral pleural involvement (%)					
No	118 (52.9)	50 (52.6)	193 (43.0)	38 (64.4)	399 (48.3)
Yes	105 (47.1)	45 (47.4)	256 (57.0)	21 (35.6)	427 (51.7)
Vascular involvement (%)					
No	209 (93.7)	90 (94.7)	443 (98.7)	49 (83.1)	791 (95.8)
Yes	14 (6.3)	5 (5.3)	6 (1.3)	8 (13.6)	33 (4.0)
Unknown	0 (0)	0 (0)	0 (0)	2 (3.4)	2 (0.2)
Type of surgery (%)					
Lobectomy or pneumonectomy	204 (91.5)	85 (89.5)	438 (97.6)	0 (0)	727 (88.0)
Wedge or segmentectomy	19 (8.5)	10 (10.5)	11 (2.4)	0 (0)	40 (4.8)
Unknown	0 (0)	0 (0)	0 (0)	59 (100)	59 (7.1)

Continued

Table 1 Continued

Characteristics	Training dataset n=223	Validation dataset n=95	Testing dataset n=449	NSCLC Radiogenomics n=59	Total n=826
EGFR mutant					
Negative	31 (13.9)	32 (33.7)	0 (0)	38 (64.4)	101 (12.2)
Positive	66 (29.6)	46 (48.4)	0 (0)	14 (23.7)	126 (15.3)
Unspecified	126 (56.5)	17 (17.9)	449 (100.0)	7 (11.9)	599 (72.5)
ALK rearrangement					
Negative	161 (72.2)	82 (86.3)	67 (14.9)	47 (79.7)	357 (43.2)
Positive	14 (6.3)	3 (3.2)	5 (1.1)	1 (1.7)	23 (2.8)
Unspecified	48 (21.5)	10 (10.5)	377 (84.0)	11 (18.6)	446 (54.0)
PD-L1					
< 1%	0 (0)	0 (0)	30 (6.7)	0 (0)	30 (3.6)
≥ 1% and < 50%	0 (0)	0 (0)	68 (15.1)	0 (0)	68 (8.2)
≥ 50%	0 (0)	0 (0)	20 (4.5)	0 (0)	20 (2.4)
Unspecified	223 (100.0)	95 (100.0)	331 (73.7)	59 (100.0)	708 (85.7)
Adjuvant systemic therapy (%)					
Only observation	174 (78.0)	72 (75.8)	381 (84.9)	41 (69.5)	668 (80.9)
Taxane +Platinum	28 (12.6)	17 (17.9)	13 (2.9)	0 (0)	58 (7.0)
Gemcitabine +Platinum	10 (4.5)	1 (1.1)	10 (2.2)	0 (0)	21 (2.5)
Vinorelbine+Platinum	5 (2.2)	0 (0)	0 (0)	0 (0)	5 (0.6)
Pemetrexed+Platinum	2 (0.9)	1 (1.1)	33 (7.3)	0 (0)	36 (4.4)
Pemetrexed+Platinum + Atezolizumab	0 (0)	1 (1.1)	1 (0.2)	0 (0)	2 (0.2)
EGFR-TKI	1 (0.4)	0 (0)	5 (1.1)	0 (0)	6 (0.7)
Unspecified therapy	3 (1.3)	3 (3.2)	6 (1.3)	18 (30.5)	30 (3.6)
Follow-up (month, mean (95% CI))†	98.3 (92.7, 103.9)	85.0 (79.5, 90.6)	27.1 (24.1, 30.0)	54.9 (46.9, 62.9)	56.3 (52.8, 59.8)
No. of DFS events	89 (39.9)	49 (51.6)	88 (19.6)	27 (45.8)	253 (30.6)

*Based on the eighth edition of the American Joint Committee on Cancer (AJCC) staging system.
†The mean follow-up was estimated using the reverse Kaplan–Meier method.
ALK, anaplastic lymphoma kinase; DFS, disease-free survival; EGFR, epidermal growth factor receptor; PD-L1, programmed cell death ligand 1; TKI, tyrosine kinase inhibitor.

were consistently higher than those of any unimodal score (figure 2B–D), underscoring its robust performance. The heatmap in figure 2E–G indicates that the unimodal scores provide complementary prognostic information, thereby enhancing the prognostic ability of the nomogram by integrating multimodal data.

Moreover, in the NSCLC Radiogenomics dataset (which includes only CT and clinical data), the radiomics plus clinical score model also outperformed the individual unimodal scores (tumor, peritumor, and clinical scores; Online supplemental figure S3), further demonstrating the benefits of multimodal integration.

The prognostic value of the multimodal nomogram was compared with that of other machine learning methods (online supplemental table S6). The integration effectiveness was found to be comparable between the multimodal nomogram (using stepwise Cox regression) and other models, including the survival support vector machine, random survival forest, and CoxBoost. The C-index for the multimodal nomogram ranged from 0.730 to 0.769, compared with 0.737 to 0.761 for the survival support vector machine, 0.724 to 0.777 for the random survival forest, and 0.729 to 0.769 for CoxBoost across the three datasets.

Stratification ability of the multimodal nomogram-defined high-/low-risk groups

Using the optimal cut-off of the multimodal nomogram, patients were classified into low-risk and high-risk groups. The low-risk group, comprising those below the cut-off value, exhibited a higher 3 year DFS rate, while the

Table 2 Uni- and multivariable Cox regression analyses for disease-free survival in the training, validation, and testing datasets

Variables	Training dataset HR (95% CI)	P	Validation dataset HR (95% CI)	P	Testing dataset HR (95% CI)	P
Univariable analysis						
Sex						
Female vs male	0.76 (0.50 to 1.16)	0.201	0.72 (0.40 to 1.29)	0.275	0.70 (0.46 to 1.07)	0.096
Smoking status						
Current or former vs never	1.33 (0.83 to 2.11)	0.232	1.92 (1.07 to 3.45)	0.028	1.25 (0.79 to 1.99)	0.336
pTNM stage						
Stage II vs stage I	2.80 (1.52 to 5.15)	0.001	1.81 (0.83 to 3.92)	0.134	2.23 (1.28 to 3.90)	0.005
Stage III vs stage I	4.41 (2.65 to 7.35)	<0.001	5.53 (2.89 to 10.58)	<0.001	5.88 (3.53 to 9.79)	<0.001
Differentiation						
Poor vs well or moderate	2.50 (1.55 to 4.02)	<0.001	3.41 (1.92 to 6.04)	<0.001	2.44 (1.59 to 3.75)	<0.001
Visceral pleural involvement						
Yes vs no	1.24 (0.81 to 1.89)	0.319	1.63 (0.93 to 2.87)	0.090	1.23 (0.80 to 1.88)	0.353
Type of surgery						
Wedge or segmentectomy vs lobectomy or pneumonectomy	1.22 (0.61 to 2.44)	0.566	0.59 (0.21 to 1.63)	0.305	1.55 (0.49 to 4.92)	0.457
Tumor score	2.72 (1.93 to 3.83)	<0.001	4.13 (2.46 to 6.94)	<0.001	3.21 (2.05 to 5.04)	<0.001
Peritumor score	2.72 (2.03 to 3.64)	<0.001	2.15 (1.65 to 2.81)	<0.001	2.47 (1.88 to 3.24)	<0.001
Immune score	2.72 (1.60 to 4.60)	<0.001	1.32 (0.70 to 2.47)	0.391	3.09 (1.64 to 5.83)	<0.001
Clinical score	2.72 (2.07 to 3.57)	<0.001	2.06 (1.45 to 2.92)	<0.001	2.31 (1.75 to 3.04)	<0.001
Multimodal nomogram	2.72 (2.14 to 3.44)	<0.001	2.21 (1.67 to 2.92)	<0.001	2.62 (2.07 to 3.33)	<0.001
Multimodal nomogram (high/low-risk groups)						
High-risk vs low-risk	5.10 (3.30 to 7.89)	<0.001	4.41 (2.40 to 8.12)	<0.001	5.54 (3.58 to 8.57)	<0.001
Multivariable analysis*						
Sex						
Female vs male	0.87 (0.50 to 1.51)	0.618	0.88 (0.41 to 1.88)	0.734	0.82 (0.47 to 1.45)	0.505
Smoking status						
Current or former vs never	1.27 (0.70 to 2.29)	0.429	0.73 (0.32 to 1.65)	0.445	0.95 (0.52 to 1.73)	0.859
pTNM stage						
Stage II vs stage I	0.75 (0.34 to 1.64)	0.466	0.61 (0.24 to 1.54)	0.291	1.24 (0.62 to 2.49)	0.546
Stage III vs stage I	0.99 (0.47 to 2.13)	0.990	1.61 (0.64 to 4.04)	0.314	1.81 (0.79 to 4.19)	0.163
Differentiation						
Poor vs well or moderate	0.89 (0.51 to 1.56)	0.688	2.85 (1.44 to 5.65)	0.003	1.64 (1.01 to 2.65)	0.046
Visceral pleural involvement						
Yes vs no	0.76 (0.47 to 1.22)	0.254	0.48 (0.23 to 1.02)	0.056	0.97 (0.59 to 1.59)	0.905
Multimodal nomogram	3.02 (2.09 to 4.35)	<0.001	2.33 (1.58 to 3.44)	<0.001	2.03 (1.36 to 3.03)	0.001
Multivariable analysis†						
Sex						
Female vs male	0.93 (0.53 to 1.63)	0.807	1.13 (0.53 to 2.45)	0.748	0.71 (0.40 to 1.25)	0.234
Smoking status						
Current or former vs never	1.40 (0.78 to 2.50)	0.257	0.73 (0.31 to 1.75)	0.482	0.93 (0.51 to 1.70)	0.820
pTNM stage						
Stage II vs stage I	0.96 (0.42 to 2.19)	0.926	0.53 (0.16 to 1.71)	0.287	1.17 (0.57 to 2.41)	0.661
Stage III vs stage I	1.64 (0.81 to 3.35)	0.170	1.76 (0.63 to 4.90)	0.276	1.85 (0.86 to 4.01)	0.117

Table 2 Continued

Variables	Training dataset		Validation dataset		Testing dataset	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Differentiation						
Poor vs well or moderate	1.23 (0.70 to 2.17)	0.475	3.44 (1.61 to 7.32)	0.001	1.82 (1.13 to 2.94)	0.014
Visceral pleural involvement						
Yes vs no	0.96 (0.60 to 1.55)	0.878	0.53 (0.23 to 1.22)	0.136	1.10 (0.68 to 1.76)	0.704
Multimodal nomogram (high/low-risk groups)						
High-risk vs low-risk	4.22 (2.26 to 7.87)	<0.001	5.01 (1.67 to 15.08)	0.004	3.59 (1.83 to 7.04)	<0.001

*The multimodal nomogram (continuous) and variables achieve p<0.20 in the univariable analysis are included in the multivariable analysis.
†The multimodal nomogram (high/low-risk grouping) and variables achieve p<0.20 in the univariable analysis are included in the multivariable analysis.

high-risk group showed a significantly lower 3 year DFS rate (figure 2E–G). In the training dataset, 158 cases (70.9%) and 57 cases (25.6%) were classified as low-risk and high-risk, respectively, with a median DFS of 169.2 months for the low-risk group and 18.7 months for the high-risk group. In the validation dataset, 58 cases (61.1%) were classified as low-risk and 30 cases (31.6%) as high-risk, with a median DFS of 27.5 months in the high-risk group, while the median DFS was not reached in the low-risk group. In the testing dataset, 305 cases (67.9%) were classified as low-risk and 132 cases (29.4%) as high-risk, with a median DFS of 83.0 months for the low-risk group and 20.8 months for the high-risk group. Across all datasets, the high-risk group defined by the multimodal nomogram consistently exhibited significantly shorter DFS after surgery compared with the low-risk group (training dataset: HR=5.10, 95% CI 3.30 to 7.89, p<0.001, figure 2H; validation dataset: HR=4.41, 95% CI 2.40 to 8.12, p<0.001, figure 2I; testing dataset: HR=5.54, 95% CI 3.58 to 8.57, p<0.001, figure 2J), demonstrating promising stratification ability.

Multivariable Cox regression adjusting for other clinicopathological risk factors was conducted to assess whether the multimodal nomogram was independently associated with DFS. The analysis confirmed an independent association between DFS and the multimodal nomogram (training dataset: HR=3.02, 95% CI 2.09 to 4.35, p<0.001; validation dataset: HR=2.33, 95% CI 1.58 to 3.44, p<0.001; testing dataset: HR=2.03, 95% CI 1.36 to 3.03, p=0.001; table 2), as well as its defined high/low-risk groups (training dataset: HR=4.22, 95% CI 2.26 to 7.87, p<0.001; validation dataset: HR=5.01, 95% CI 1.67 to 15.08, p=0.004; testing dataset: HR=3.59, 95% CI 1.83 to 7.04, p<0.001; table 2).

Subgroup analysis was further performed on subgroups stratified by clinicopathological risk factors. Kaplan–Meier curves demonstrated that the multimodal nomogram-defined high-/low-risk groups exhibited robust prognostic stratification ability across all subgroups (p<0.05, figure 3).

Predictive value of the multimodal nomogram-defined adjuvant chemotherapy benefit rate

To guide individualized adjuvant therapy decisions and identify patients who may benefit from adjuvant chemotherapy, two additional multimodal nomograms were constructed using data from the treatment arms in the training dataset: the observation nomogram (figure 4A) and the adjuvant chemotherapy nomogram (figure 4B). The difference in the 3 year DFS rate predicted by these two nomograms was used to calculate the adjuvant chemotherapy benefit rate. Patients were then stratified into three groups based on this estimated benefit rate: ≤10%, >10% and ≤25%, and >25%. Patients receiving targeted therapy (n=6), immunotherapy (n=2), and unspecified therapy (n=12) were excluded from this analysis.

DFS difference between treatment arms across different benefit rate groups was assessed in the pooled dataset (stages I–III). Patients with a benefit rate of ≤10% had longer DFS without adjuvant chemotherapy (HR=2.20, 95% CI 1.31 to 3.69, p=0.003, figure 4C). Patients with a benefit rate of >10% and ≤25% had similar DFS between the two treatment arms (HR=0.33, 95% CI 0.07 to 1.47, p=0.145, figure 4D), while patients with a benefit rate of >25% had longer DFS with adjuvant chemotherapy (HR=0.01, 95% CI <0.001 to 0.51, p=0.025, figure 4E). After adjusting for the treatment main effect, benefit rate main effect, and other clinicopathological risk factors, there was a significant interaction between treatment arms and benefit rate group in the pooled dataset (adjusted interaction p<0.001), indicating that the multimodal nomogram-defined adjuvant chemotherapy benefit rate could serve as a predictive biomarker.

Further analysis was conducted on candidates for adjuvant chemotherapy identified by the ESMO[6] and NCCN[5] guidelines (stage IB–IIA patients with high-risk features for recurrence or tumor >4 cm). In this subgroup, patients with a benefit rate of ≤10% had longer DFS without adjuvant chemotherapy (HR=2.05, 95% CI 1.50 to 2.80, p<0.001, figure 4F), while those with a benefit rate of >10% and ≤25% showed a trend
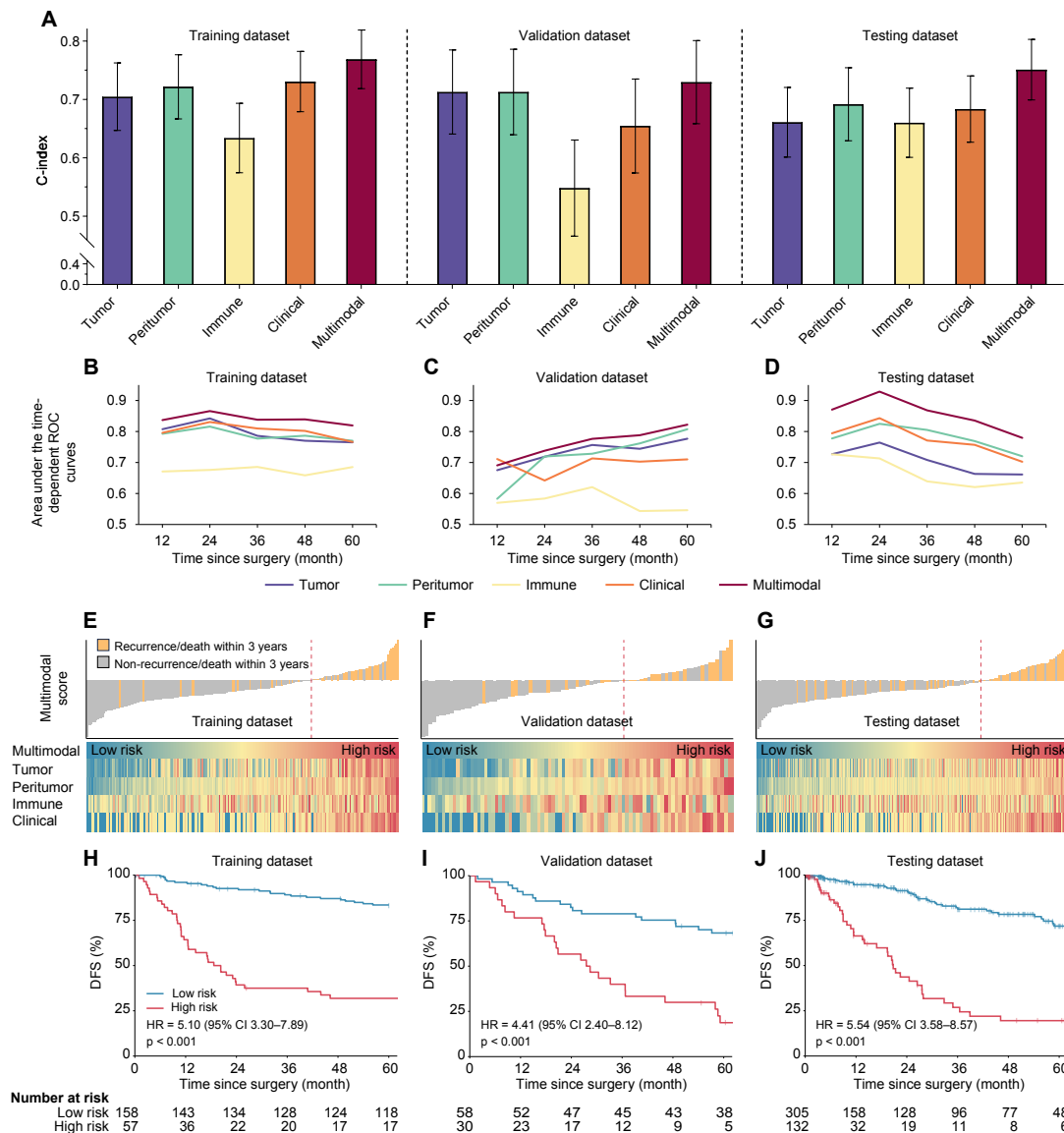
Figure 2 Prognostic performance comparison of unimodal scores and multimodal nomogram. (A) The C-index and 95% CIs for disease-free survival (DFS) prediction using the unimodal scores and the multimodal nomogram. Area under the time-dependent ROC curves for predicting 1, 2, 3, 4, and 5 year DFS using the unimodal scores and the multimodal nomogram in the training (B), validation (C), and testing datasets (D). (E–G) Waterfall plots demonstrate the incidence of DFS events as the multimodal nomogram score increases from left to right, with the high-risk group to the right of the red dashed line showing a significantly higher incidence of DFS events compared with the low-risk group on the left. Heatmaps display that the tumor, peritumor, immune, and clinical scores provide complementary prognostic information. (H–J) Kaplan–Meier curves show that the multimodal nomogram-defined high/low-risk groups have promising stratification ability across multicenter datasets.

of prolonged DFS with adjuvant chemotherapy, although not significant (HR=1.06, 95% CI 0.49 to 2.80, p=0.883, figure 4G), suggesting that the adjuvant chemotherapy benefit rate may support individualized treatment decisions for candidates for adjuvant chemotherapy.

Additionally, in candidates for adjuvant chemotherapy, the multimodal nomogram-defined high-/low-risk groups could further refine the stratification in this subgroup. This was validated in Center A (n=104, HR=3.88, 95% CI 2.07 to 7.28, p<0.001, figure 4H) and Center B (n=188, HR=4.86, 95% CI 2.14 to 11.01, p<0.001, figure 4I). Low-risk candidates exhibited survival curves similar to patients who do not require

adjuvant chemotherapy according to the guidelines (stage I patients without high-risk features), while high-risk candidates exhibited survival curves similar to patients requiring adjuvant chemotherapy (stage IIB–III).

Correlation between immune score and immune phenotype

The immune phenotype has been widely recognized in profiling the immune microenvironment of LUAD in recent years. We explored its correlation with immune score to understand the value of immune score across different immune landscapes. The Pearson χ^2 test revealed statistically significant differences in immune score across the inflamed, immune-excluded, and immune-desert
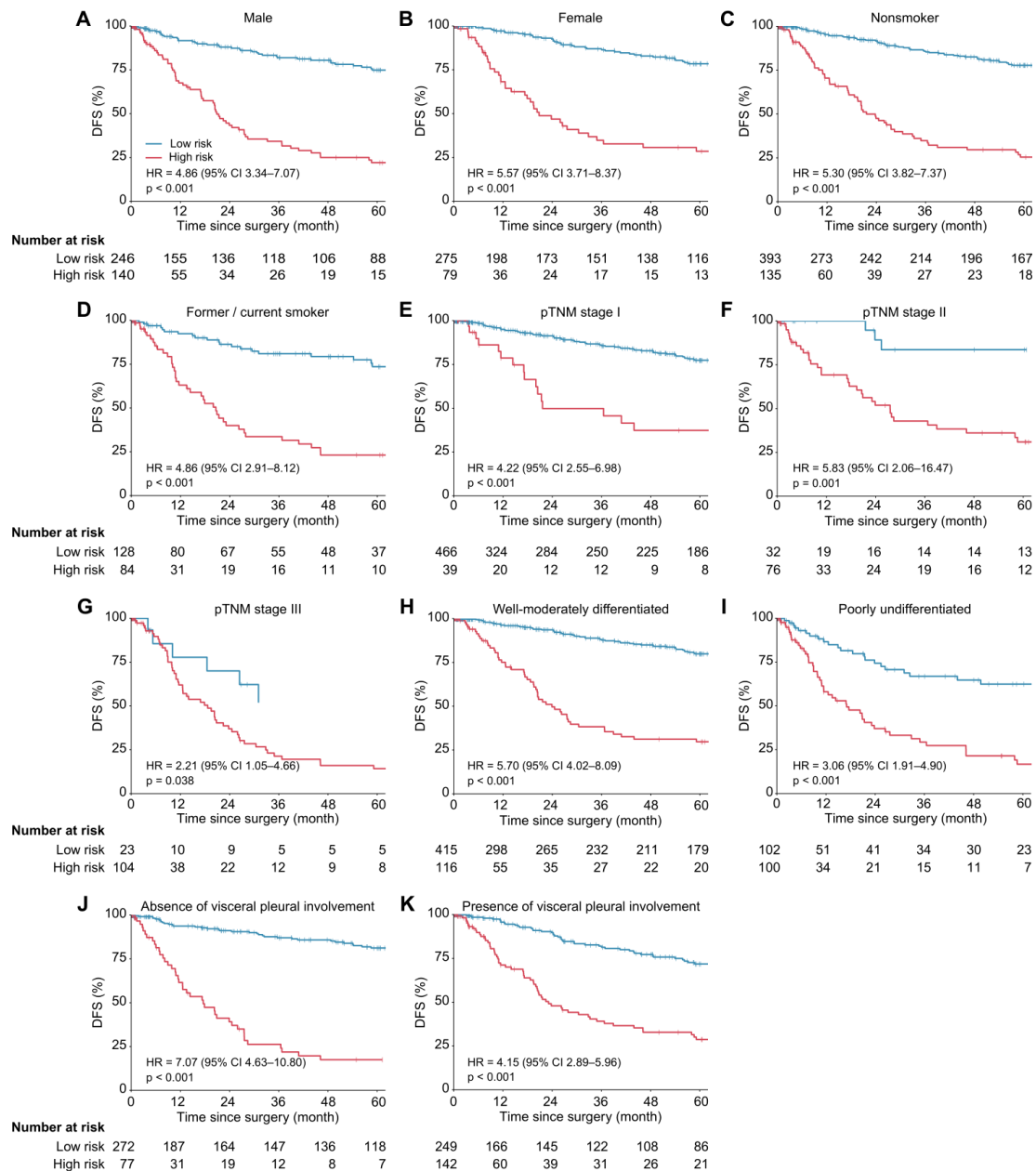
Figure 3 Kaplan–Meier curves for subgroup analysis. After pooling the training, validation, and testing datasets, the prognostic ability of the multimodal nomogram-defined high-/low-risk groups was verified in the subgroups stratified by clinicopathological risk factors, including (A) males, (B) females, (C) non-smokers, (D) former/current smokers, (E) stage I disease, (F) stage II disease, (G) stage III disease, (H) well-moderately differentiated tumors, (I) poorly differentiated tumors, (J) absence of visceral pleural involvement, and (K) presence of visceral pleural involvement.

phenotypes (p<0.001 in the training validation, testing, and pooled dataset). Among the three immune phenotypes, the inflamed phenotype had the lowest immune score and the immune-desert phenotype had the highest immune score (figure 5I).

In the pooled dataset, the prognostic value of the immune score was assessed in the subgroups of three immune phenotypes. Cox regression analysis demonstrated that the immune score remained significantly associated with DFS within the immune-desert phenotype (HR=3.68, 95% CI 1.52 to 8.91, p=0.004). However, the association between immune score and DFS did not

reach statistical significance in the immune-excluded phenotype (HR=2.29, 95% CI 0.81 to 6.49, p=0.117) or the inflamed phenotype (HR=1.21, 95% CI 0.67 to 2.17, p=0.535).

Histopathological and biological basis of the tumor and peritumor scores

The radiomics-based tumor and peritumor scores were derived from macroscopic CT imaging, with underlying mechanisms remaining unclear. Therefore, we explored their histopathological basis by analyzing their association
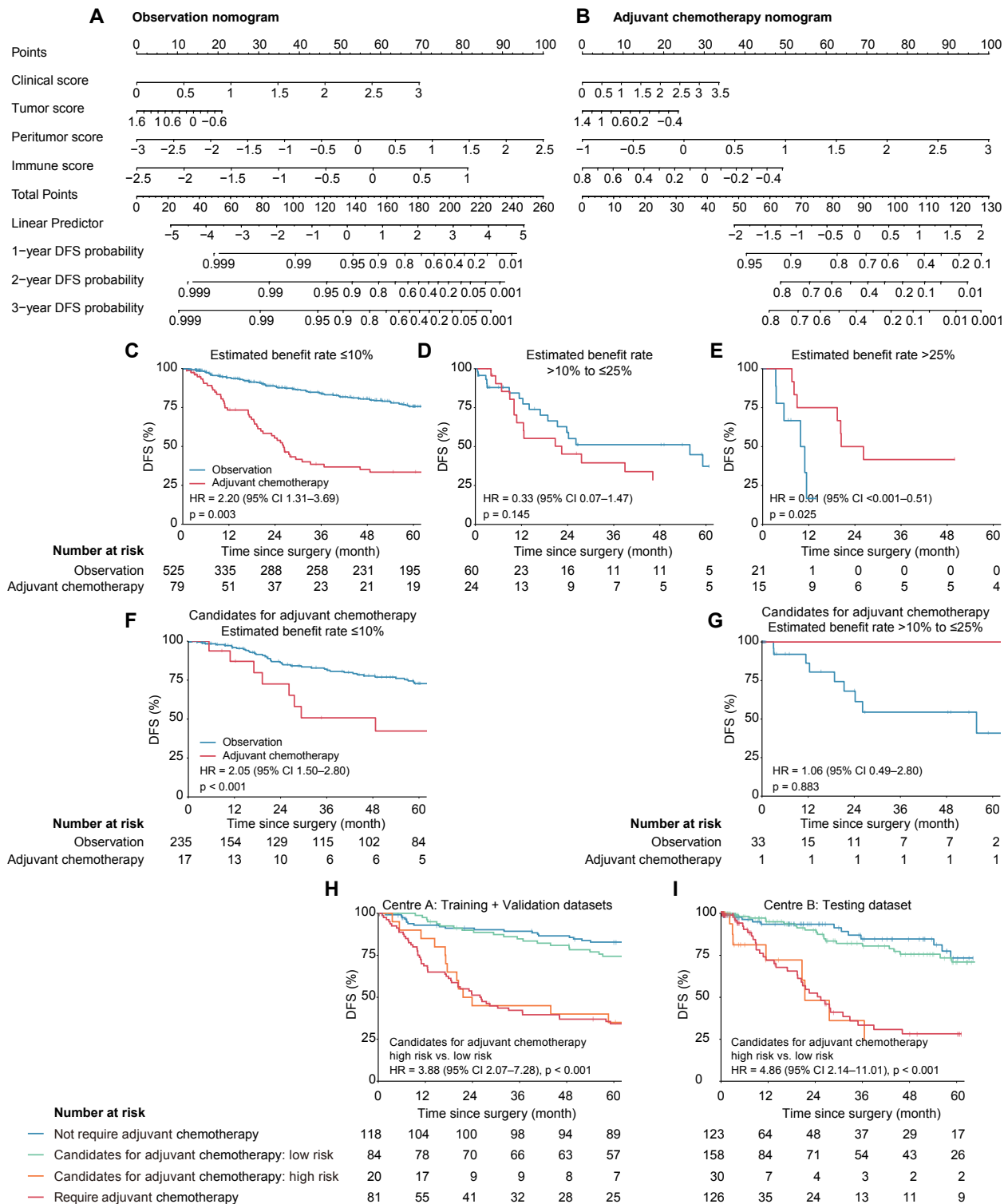
Figure 4 Predictive value of the multimodal nomogram-defined adjuvant chemotherapy benefit. (A) The observation nomogram and (B) the adjuvant chemotherapy nomogram. The estimated adjuvant chemotherapy benefit rate is calculated as the difference between the 3 year disease-free survival (DFS) rate predicted by these two nomograms. (C) Patients with a benefit rate of ≤10% have longer DFS without adjuvant chemotherapy. (D) Patients with a benefit rate of >10% and ≤25% have similar DFS between the two treatment arms. (E) Patients with a benefit rate of >25% have longer DFS with adjuvant chemotherapy. In candidates for adjuvant chemotherapy identified by the European Society for Medical Oncology and National Comprehensive Cancer Network guidelines, (F) patients with a benefit rate of ≤10% have longer DFS without adjuvant chemotherapy, while (G) those with a benefit rate of >10% and ≤25% show a trend of prolonged DFS with adjuvant chemotherapy. In Center A (H) and Center B (I), among candidates for adjuvant chemotherapy, the multimodal nomogram-defined high-/low-risk groups can significantly distinguish between patients with superior (green lines) and poorer (orange lines) outcomes, thereby guiding individualized treatment decisions.
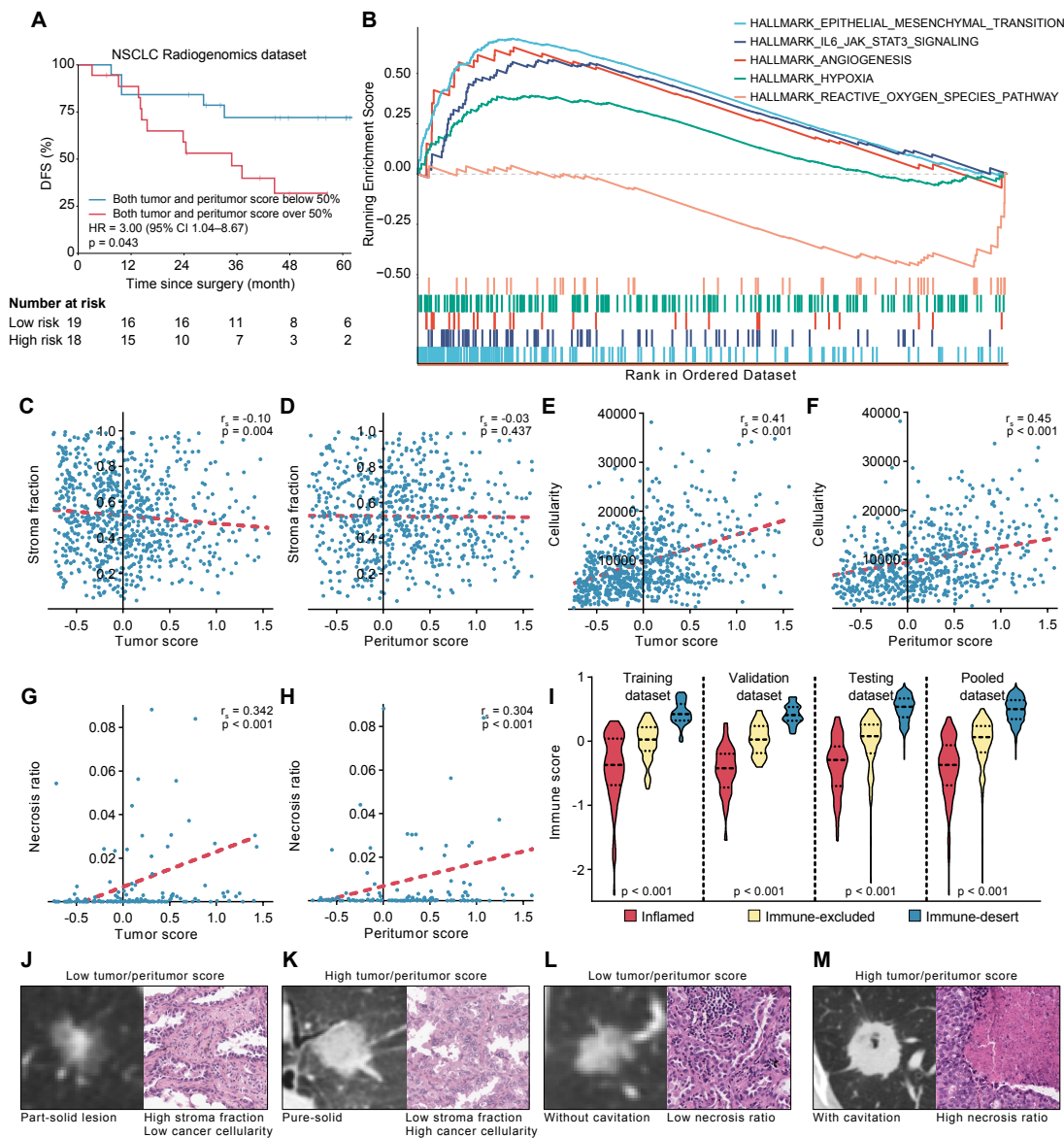
Figure 5 Histopathological and biological basis of the tumor, peritumor, and immune scores. (A) Kaplan–Meier curves for the high-risk group (upper 50% in both tumor and peritumor scores) and the low-risk group (lower 50% in both scores) in the NSCLC Radiogenomics dataset. (B) Gene set enrichment analysis results. (**C–H**) Cross-modal correlation analysis between the radiomics-based tumor/peritumor scores and histopathological quantitative metrics. (I) Correlation between immune score and immune phenotype. (**J–M**) CT images and histopathological images (×100 magnification) of typical cases with low or high tumor/peritumor scores.

with histopathological quantitative metrics and their biological basis through GSEA.

Spearman's correlation analysis revealed a negative correlation between the tumor score and stroma fraction (r_s=−0.10, p=0.004, figure 5C), but no statistically significant correlation was found between the peritumor score and stroma fraction (r_s=−0.03, p=0.437, figure 5D). Both the tumor score (r_s=0.41, p<0.001, figure 5E) and peritumor score (r_s=0.45, p<0.001, figure 5F) showed positive correlations with cancer cellularity. Cases with low tumor/peritumor scores typically appeared as part-solid lesions on CT, with a high stroma fraction and low cancer cellularity on histopathological images (figure 5J). In contrast, cases with high tumor/peritumor scores typically

appeared as pure-solid lesions on CT, with a low stroma fraction and high cancer cellularity on histopathological images (figure 5K). Both the tumor score (r_s=0.342, p<0.001, figure 5G) and peritumor score (r_s=0.304, p<0.001, figure 5H) were positively correlated with necrosis ratio. Cases with low tumor/peritumor scores were characterized by no cavitation on CT and a low necrosis ratio on histopathological images (figure 5L), while those with high tumor/peritumor scores showed cavitation formation on CT and a high necrosis ratio on histopathological images (figure 5M).

In the NSCLC Radiogenomics dataset, 18 cases were classified as high-risk (upper 50% in both tumor and peritumor scores) and 19 as low-risk (lower 50% in

both scores). Kaplan–Meier curves demonstrated that low-risk patients had superior DFS compared with high-risk patients (HR=3.00, 95% CI 1.04 to 8.67, p=0.043, figure 5A). Subsequently, the differential expression genes between the high- and low-risk groups were subjected to GSEA, with the enriched gene sets presented in figure 5B. The detailed GSEA results are listed in online supplemental table S7. In the high-risk group, gene sets related to epithelial-mesenchymal transition (HALLMARK_EPITHELIAL_MESENCHYMAL_TRAN-SITION, NES=2.653, FDR<0.001), IL6-JAK-STAT3 signaling (HALLMARK_IL6_JAK_STAT3_SIGNALING, NES=1.971, FDR<0.001), angiogenesis (HALLMARK_ANGIOGENESIS, NES=1.812, FDR=0.003), and hypoxia (HALLMARK_HYPOXIA, NES=1.522, FDR=0.005) were significantly enriched, indicating upregulation of genes associated with these pathways as the tumor/peritumor score increased. In the low-risk group, the gene sets related to the reactive oxygen species pathway (HALL-MARK_REACTIVE_OXYGEN_SPECIES_PATHWAY, NES=−1.565, FDR=0.021) were significantly enriched, suggesting upregulation of genes associated with this pathway as the tumor/peritumor score decreased.

Interleukin-6 (IL-6) is a cytokine that regulates immune and inflammatory responses, while signal transducer and activator of transcription 3 (STAT3) serve as a crucial cellular signaling protein, which can be activated by the IL-6/gp130 receptor family, Janus kinases (JAK), etc. In approximately 50% of lung cancers, STAT3 is persistently activated, influencing cancer initiation and progression through the modulation of cell cycle, invasion, metastasis, immune evasion, and angiogenesis.[41] The activation of the STAT3 pathway also induces epithelial-mesenchymal transition, promoting invasion and metastasis of lung cancer.[42] Epithelial-mesenchymal transition has been correlated with prognosis in various malignancies.[43] Oxidative stress plays a dual role in promoting and inhibiting tumorigenesis.[44] In terms of promotion, hypoxia, a characteristic of tumor microenvironment, facilitates cancer progression.[42] In terms of inhibition, oxidative stress is crucial for the differentiation, maturation, and activation of antitumor immune cells, and high levels of reactive oxygen species can induce cancer cell death.[45]

DISCUSSION

This study constructed a multimodal nomogram by incorporating tumor score, peritumor score, immune score, and clinical score for prognosis and postoperative treatment decision in resectable LUAD. The multimodal nomogram demonstrates robust and promising prognostic performance in multicenter datasets. It can identify patients requiring adjuvant chemotherapy and those who do not and provide refined risk stratification for candidates of adjuvant chemotherapy as recommended by the NCCN and ESMO guidelines. To our knowledge, this is the first multimodal study incorporating CT imaging and immune status for LUAD prognostication.

Previous radiomics studies have emphasized the importance of peritumoral features in lung cancer prognostication.[11 46] This study has further validated the benefit of incorporating tumor, peritumor, and clinical information in the publicly available NSCLC Radiogenomics dataset. The results indicate that the radiomics plus clinical score model outperforms the unimodal scores at almost all time points, consistent with previous findings.[11 46] On this basis, we further explore the prognostic mechanisms of the tumor/peritumor scores from histopathological and biological perspectives. Histologically, cases with higher tumor/peritumor scores typically exhibit pure solid lesions and are associated with lower stroma fraction and higher cancer cellularity—radiological[47] and histological factors[34 36] that have been confirmed to be relevant to lung cancer outcomes. Moreover, higher tumor/peritumoral scores are associated with a higher necrosis ratio. Such lesions are more prone to cavity formation on CT and display extensive necrotic areas on WSIs. The presence of necrosis has been proven to be associated with poor prognosis in lung cancer.[35 48] From the biological perspective, tumor/peritumor scores are associated with epithelial-mesenchymal transition, angiogenesis, IL6-JAK-STAT3 signaling, and reactive oxidative species pathway, which are involved in the progression and outcomes of various malignancies.[41–45] There is some correspondence between these gene pathways and histopathological quantitative metrics mentioned above (eg, stroma fraction/cancer cellularity with the epithelial-mesenchymal transition; necrosis ratio with hypoxia/cancer cell death, etc), enhancing our understanding of cross-scale correlations among radiology, histopathology, and genomics.

Several studies have confirmed that apart from the density of TILs, their spatial arrangement is also closely related to anti-tumor immunity.[49 50] H&E-stained slides have an advantage in analyzing both the density and spatial arrangement of TILs in a cost-effective manner. Currently, many studies integrate the density of TILs in cancer epithelium and stroma by linear weighting[22] or calculating the proportion of grid classification,[15] then constructing prognostic models for lung cancer. This study constructed an immune score using a Cox regression model to incorporate the density of TILs in cancer epithelium and stroma. Its prognostic capability was validated in multicenter data (C-index 0.55–0.66), with performance comparable to previous results.[22] On the other hand, the immune phenotype, also based on the density of TILs, is often used in assessing the immune contexture of solid tumors.[51] The immune phenotype is typically classified as inflamed, immune-excluded, and immune-desert in lung cancer. In this study, the immune score is not only associated with the immune phenotype but also remains correlated with DFS in the subgroup of immune-desert phenotype. This indicates that the immune score can offer refined prognostic information within the framework of immune phenotype. Although it exhibits stable performance, its prognostic capability remains limited, possibly because the immune score only

depicts the antitumor immunity to some extent. Additionally, the tissue sections are obtained from local surgical specimens, thus limiting their representativeness for the entire tumor.

Conceptually, the CT-based scores can describe the macroscopic morphology of tumor and peritumor regions, while the WSI-based immune score can reflect anti-tumor immunity at the microscopic level. They provide complementary information across different scales and dimensions. In terms of prognostic capability, the multimodal nomogram demonstrates superior discrimination and calibration performance compared with any unimodal scores. Compared with previous studies on radiology-clinical integration,[11] pathology-genomics integration,[24] and other lung cancer prognosis research, our multimodal nomogram achieves optimal prognostic performance (C-index 0.751 vs 0.60–0.74 in external testing datasets). In terms of clinical utility, we defined the adjuvant chemotherapy benefit rate based on the multimodal nomogram and used it to guide the selection of patients suitable for observation or chemotherapy after surgery, avoiding undertreatment or overtreatment. In this study, 35.4% (292/826) of patients with resectable LUAD were candidates for adjuvant chemotherapy. These patients exhibit high heterogeneity, having stage IB–IIA disease with at least one high-risk feature for recurrence. Our multimodal nomogram can provide more refined risk stratification for this heterogeneous subgroup of patients. Low-risk candidates demonstrate similar DFS to those who do not require adjuvant chemotherapy (stage I disease without high-risk features for recurrence), while high-risk candidates demonstrate similar DFS to those who require adjuvant chemotherapy (stage IIB–III disease), providing additional information for clinical decision-making after surgery.

This study has the following limitations: first, as a retrospective study, the generalizability of the multimodal nomogram and its defined benefit rate for adjuvant chemotherapy requires validation in larger, prospective datasets, especially in candidates for adjuvant chemotherapy. Since all cases in this study underwent surgery before 2017, the majority of adjuvant treatment is limited to platinum-based chemotherapy. As a result, the predictive value of the multimodal nomogram has not been validated in the populations receiving targeted therapy or immunotherapy. These emerging therapies, especially immunotherapies, which are closely related to the density of TILs, should be considered for inclusion in future research. Second, TILs in the cancer epithelium and stroma compartments are identified using Hover-Net. We will continue to monitor advancements in segmentation and classification algorithms and aim to apply improved approaches to TIL quantification. Third, the immune score constructed in this study is based on the density of TILs in the cancer epithelium and stroma compartments, which characterize the spatial arrangement of TILs solely at the tissue level. Incorporating the spatial arrangement of TILs at the cellular level into the immune score could

provide additional prognostic information. Additionally, other immune microenvironment biomarkers, such as the density of CD8$^+$ T lymphocytes, PD-L1 tumor cell proportion score, and the density of tertiary lymphoid structures should be incorporated to optimize the characterization of antitumor immunity. Fourth, only patients with both contrast-enhanced CT images and H&E-stained WSIs are included in the analysis, which limits the applicability to patients who may lack one of the modalities. However, we are developing deep learning-based multimodal integration approaches that can tolerate missing modalities, enhancing the applicability in broader clinical settings.

In conclusion, the multimodal nomogram comprehensively characterizes LUAD across macroscopic and microscopic scales by integrating tumor/peritumor morphology, immune status, and clinical information. It demonstrates robust and promising prognostic capability, consistently outperforming unimodal scores. The multimodal nomogram-defined adjuvant chemotherapy benefit rate enables more refined risk stratification for candidates of adjuvant chemotherapy, thereby supporting individualized postoperative treatment decisions.

Author affiliations
[1]Department of Radiology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China
[2]Department of Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong, China
[3]School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, Guangxi, China
[4]Department of Pathology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, China
[5]Department of Radiology, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China
[6]Institute of Computational Science and Technology, Guangzhou University, Guangzhou, Guangdong, China
[7]Department of Radiology, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China
[8]Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou, Guangdong, China

X Huan Lin @huan33270199

ORCID iDs
Huan Lin http://orcid.org/0000-0003-4354-6110
Junjie Hua http://orcid.org/0000-0002-5258-4816

REFERENCES

1. Bray F, Laversanne M, Sung H, *et al*. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74:229–63.
2. Thai AA, Solomon BJ, Sequist LV, *et al*. Lung cancer. *Lancet* 2021;398:535–54.
3. Muthusamy B, Patil PD, Pennell NA. Perioperative Systemic Therapy for Resectable Non–Small Cell Lung Cancer. *J Natl Compr Canc Netw* 2022;20:953–61.
4. Rami-Porta R, Asamura H, Travis WD, *et al*. Lung cancer - major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017;67:138–55.
5. National comprehensive cancer network (nccn) guideline: non-small cell lung cancer, version 3. 2025. 2025. Available: https://www.nccn.org/ [Accessed 25 Jan 2025].
6. Remon J, Soria J-C, Peters S, *et al*. Early and locally advanced non-small-cell lung cancer: an update of the ESMO Clinical Practice Guidelines focusing on diagnosis, staging, systemic and local therapy. *Ann Oncol* 2021;32:1637–42.
7. Goldstraw P, Chansky K, Crowley J, *et al*. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 2016;11:39–51.
8. Pignon J-P, Tribodet H, Scagliotti GV, *et al*. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. *J Clin Oncol* 2008;26:3552–9.
9. Park S, Lee SM, Choe J, *et al*. Recurrence Patterns and Patient Outcomes in Resected Lung Adenocarcinoma Differ according to Ground-Glass Opacity at CT. *Radiology* 2023;307:e222422.
10. Kim SK, Kim TJ, Chung MJ, *et al*. Lung Adenocarcinoma: CT Features Associated with Spread through Air Spaces. *Radiology* 2018;289:831–40.
11. Vaidya P, Bera K, Gupta A, *et al*. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multicohort study for outcome prediction. *Lancet Digit Health* 2020;2:e116–28.
12. Zhong Y, She Y, Deng J, *et al*. Deep Learning for Prediction of N2 Metastasis and Survival for Clinical Stage I Non-Small Cell Lung Cancer. *Radiology* 2022;302:200–11.
13. Bremnes RM, Busund L-T, Kilvær TL, *et al*. The Role of Tumor-Infiltrating Lymphocytes in Development, Progression, and Prognosis of Non-Small Cell Lung Cancer. *J Thorac Oncol* 2016;11:789–800.
14. Brambilla E, Le Teuff G, Marguet S, *et al*. Prognostic Effect of Tumor Lymphocytic Infiltration in Resectable Non-Small-Cell Lung Cancer. *J Clin Oncol* 2016;34:1223–30.
15. Park S, Ock C-Y, Kim H, *et al*. Artificial Intelligence-Powered Spatial Analysis of Tumor-Infiltrating Lymphocytes as Complementary Biomarker for Immune Checkpoint Inhibition in Non-Small-Cell Lung Cancer. *J Clin Oncol* 2022;40:1916–28.
16. Donnem T, Kilvaer TK, Andersen S, *et al*. Strategies for clinical implementation of TNM-Immunoscore in resected nonsmall-cell lung cancer. *Ann Oncol* 2016;27:225–32.
17. Hu C, Shu L, Chen C, *et al*. A prediction model integrated genomic alterations and immune signatures of tumor immune microenvironment for early recurrence of stage I NSCLC after curative resection. *Transl Lung Cancer Res* 2022;11:24–42.
18. Donnem T, Hald SM, Paulsen E-E, *et al*. Stromal CD8+ T-cell Density—A Promising Supplement to TNM Staging in Non-Small Cell Lung Cancer. *Clin Cancer Res* 2015;21:2635–43.
19. Al-Shibli KI, Donnem T, Al-Saad S, *et al*. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clin Cancer Res* 2008;14:5220–7.
20. Rakaee M, Adib E, Ricciuti B, *et al*. Association of Machine Learning–Based Assessment of Tumor-Infiltrating Lymphocytes on Standard Histologic Images With Outcomes of Immunotherapy in Patients With NSCLC. *JAMA Oncol* 2023;9:51.
21. Lin H, Pan X, Feng Z, *et al*. Automated whole-slide images assessment of immune infiltration in resected non-small-cell lung cancer: towards better risk-stratification. *J Transl Med* 2022;20:261.
22. Pan X, Lin H, Han C, *et al*. Computerized tumor-infiltrating lymphocytes density score predicts survival of patients with resectable lung adenocarcinoma. *iScience* 2022;25:105605.
23. Gui C-P, Chen Y-H, Zhao H-W, *et al*. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. *The Lancet Digital Health* 2023;5:e515–24.
24. Chen RJ, Lu MY, Williamson DFK, *et al*. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* 2022;40:865–78.
25. Feng L, Liu Z, Li C, *et al*. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health* 2022;4:e8–17.
26. Boehm KM, Aherne EA, Ellenson L, *et al*. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat Cancer* 2022;3:723–33.
27. Vanguri RS, Luo J, Aukerman AT, *et al*. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer* 2022;3:1151–64.
28. Boehm KM, Khosravi P, Vanguri R, *et al*. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer* 2022;22:114–26.
29. Lipkova J, Chen RJ, Chen B, *et al*. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* 2022;40:1095–110.
30. Isensee F, Jaeger PF, Kohl SAA, *et al*. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* 2021;18:203–11.
31. Feng Z, Lin H, Liu Z, *et al*. Artificial intelligence-quantified tumour-lymphocyte spatial interaction predicts disease-free survival in resected lung adenocarcinoma: A graph-based, multicentre study. *Comput Methods Programs Biomed* 2023;238:107617.
32. Han C, Lin J, Mai J, *et al*. Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. *Med Image Anal* 2022;80:102487.
33. Graham S, Vu QD, Raza SEA, *et al*. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* 2019;58:101563.
34. Micke P, Strell C, Mattsson J, *et al*. The prognostic impact of the tumour stroma fraction: A machine learning-based analysis in 16 human solid tumour types. *EBioMedicine* 2021;65:103269.
35. Oiwa H, Aokage K, Suzuki A, *et al*. Clinicopathological, gene expression and genetic features of stage I lung adenocarcinoma with necrosis. *Lung Cancer (Auckl)* 2021;159:74–83.

36 Lin S, Samsoondar JP, Bandari E, *et al*. Digital Quantification of Tumor Cellularity as a Novel Prognostic Feature of Non-Small Cell Lung Carcinoma. *Mod Pathol* 2023;36:100055.

37 Park J, Cho H-G, Park J, *et al*. Artificial Intelligence-Powered Hematoxylin and Eosin Analyzer Reveals Distinct Immunologic and Mutational Profiles among Immune Phenotypes in Non-Small-Cell Lung Cancer. *Am J Pathol* 2022;192:701–11.

38 Wu T, Hu E, Xu S, *et al*. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;2:100141.

39 Liberzon A, Birger C, Thorvaldsdóttir H, *et al*. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–25.

40 Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004;10:7252–9.

41 Gao SP, Mark KG, Leslie K, *et al*. Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. *J Clin Invest* 2007;117:3846–56.

42 Zhang X, Sai B, Wang F, *et al*. Hypoxic BMSC-derived exosomal miRNAs promote metastasis of lung cancer cells via STAT3-induced EMT. *Mol Cancer* 2019;18:40.

43 De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* 2013;13:97–110.

44 Li K, Deng Z, Lei C, *et al*. The Role of Oxidative Stress in Tumorigenesis and Progression. *Cells* 2024;13:441.

45 Gorrini C, Harris IS, Mak TW. Modulation of oxidative stress as an anticancer strategy. *Nat Rev Drug Discov* 2013;12:931–47.

46 Liu K, Li K, Wu T, *et al*. Improving the accuracy of prognosis for clinical stage I solid lung adenocarcinoma by radiomics models covering tumor per se and peritumoral changes on CT. *Eur Radiol* 2022;32:1065–77.

47 Ye T, Deng L, Wang S, *et al*. Lung Adenocarcinomas Manifesting as Radiological Part-Solid Nodules Define a Special Clinical Subtype. *J Thorac Oncol* 2019;14:617–27.

48 Swinson DEB, Jones JL, Richardson D, *et al*. Tumour necrosis is an independent prognostic marker in non-small cell lung cancer: correlation with biological variables. *Lung Cancer* 2002;37:235–40.

49 Barua S, Fang P, Sharma A, *et al*. Spatial interaction of tumor cells and regulatory T cells correlates with survival in non-small cell lung cancer. *Lung Cancer (Auckl)* 2018;117:73–9.

50 Galon J, Costes A, Sanchez-Cabo F, *et al*. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006;313:1960–4.

51 Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov* 2019;18:197–218.