RESEARCH

Habitat radiomics analysis for progression free survival and immune-related adverse reaction prediction in non-small cell lung cancer treated by immunotherapy

Yuemin Wu[1], Wei Zhang[2], Xiao Liang[1], Pengpeng Zhang[3], Mengzhe Zhang[3], Yuqin Jiang[1], Yanan Cui[2], Yi Chen[4], Wenxin Zhou[1], Qi Liang[1], Jiali Dai[1], Chen Zhang[1], Jiali Xu[1], Jun Li[1], Tongfu Yu[2], Zhihong Zhang[5] and Renhua Guo[2*]

Abstract

Background Non-small cell lung cancer (NSCLC) is highly heterogeneous, leading to varied treatment responses and immune-related adverse reactions (irAEs) among patients. Habitat radiomics allows non-invasive quantitative assessment of intratumor heterogeneity (ITH). Therefore, our objective is to employ habitat radiomics techniques to develop a robust approach for predicting the efficacy of Immune checkpoint inhibitors (ICIs) and the likelihood of irAEs in advanced NSCLC patients.

Methods In this retrospective two center study, two independent cohorts of patients with NSCLC were used to develop (n = 248) and validate signatures (n = 95). After applying four kinds of machine learning algorithms to select the key preoperative CT radiomic features, we used clinical, radiomics and habitat radiomic features to develop the clinical signature, radiomics signature and habitat radiomic signature for ICIs prognostics and irAEs prediction. By combining habitat radiomic features with corresponding clinicopathologic information, the nomogram signature was constructed in the training cohort. Next, the internal validation cohort (n = 75) of patients, and the external validation cohort (n = 20) of patients treated with ICIs were included to evaluate the predictive value of the four signatures, and their predictive performance was assessed by the area under operating characteristic curve (AUC).

Results Our study introduces a radiomic nomogram model that integrates clinical and habitat radiomic features to identify patients who may benefit from ICIs or experience irAEs. The Radiomics Nomogram model exhibited superior predictive performance in the training, validation, and external validation sets, with AUCs of 0.923, 0.817, and 0.899, respectively. This model outperformed both the Whole-tumor Radiomics Signature model (AUCs of 0.870, 0.736, and 0.626) and the Habitat Signature model (AUCs of 0.900, 0.804, and 0.808). The radiomics model focusing on tumor sub-regional habitat showed better predictive performance than the model derived from the entire tumor. Decision Curve Analysis (DCA) and calibration curves confirmed the nomogram's effectiveness.

*Correspondence:
Renhua Guo
rhguo@njmu.edu.cn
Full list of author information is available at the end of the article

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 2 of 19

Conclusion By leveraging machine learning to predict the outcomes of ICIs, we can move closer to achieving tailored ICIs for lung cancer. This advancement will assist physicians in selecting and managing subsequent treatment strategies, thereby facilitating clinical decision-making.

Keywords Non-small cell lung cancer, Machine learning, Habitat radiomics, Tumor microenvironment, Immune-related adverse events

Introduction

Lung cancer stands as the most prevalent malignant tumor and the primary cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) representing 85% of lung cancer cases [1, 2].

In recent years, the extensive use of Immunotherapy (IO), notably Immune checkpoint inhibitors (ICIs), has transformed the treatment landscape for various cancer types [3, 4]. Despite the promising outcomes of ICIs, not all patients benefit equally from ICIs. Response rates to ICIs can vary depending on patients' distinct molecular, histological, or genetic characteristics.

Moreover, the effective implementation of ICIs faces obstacles due to immune-related adverse events (irAEs), presenting significant challenges. irAEs are not uncommon during clinical treatment. The utilization of ICIs in cancer therapy has led to a substantial increase in the yearly incidence of irAEs, with cases escalating dramatically [5]. By 2018, nearly 13,000 cases had been reported. A comprehensive analysis of 13,051 cases was conducted using Big Data Analysis. A meta-analysis encompassing 36 phase II/III trials, evaluating the overall safety of ICIs, revealed a pooled incidence rate ranging from 54 to 76% for all adverse events [6]. In the as-treated population of the KEYNOTE-407 study, irAEs and infusion reactions occurred in 35.6% of patients (grade 3–5, 13.3%) and 9.3% of patients (grade 3–5, 3.2%), respectively [7].

Although severe irAEs are rare, specific immune responses can cause significant discomfort and pose life-threatening risks to patients. Consequently, early detection plays a crucial role in ensuring optimal patient care. Distinguishing between patients' responses to ICIs and assessing varying risks of irAEs has become paramount in tailoring individualized treatment strategies.

Tumor heterogeneity poses a persistent challenge in the realm of cancer therapy, manifesting in variations between cancers from different patients (inter-tumor heterogeneity) and within individual tumors (intra-tumor heterogeneity (ITH)). The latter encompasses phenotypic diversity, such as cell surface markers, (epi)genetic abnormalities, growth rates, apoptosis, and other cancer hallmarks that drive disease progression and treatment resistance [8]. This heterogeneity may contribute to disparate responses to ICIs in NSCLC, particularly among patients with similar clinical characteristics (e.g., clinical stage, molecular subtype) [7, 8]. Consequently, tumor biomarkers reflecting characteristics of the entire lung tumor while overlooking ITH have demonstrated limited predictive value for treatment response [9]. Conventional Radiomics has obscured intra-tumor complexity, hindering a comprehensive exploration of the impact of ITH on treatment efficacy.

ITH significantly influences the effectiveness of various immunotherapies and complicates the identification of clinically prognostic biomarkers. Therefore, a quantitative assessment of ITH could serve as a valuable biomarker for predicting clinical outcomes [9, 10].

Artificial intelligence, as an emerging field, has showcased remarkable capabilities across various domains, with particular attention in the medical and health sectors due to its potential to enhance diverse aspects of medical services, including diagnosis, risk analysis, lifestyle monitoring, and more [11, 12].

In the field of medical imaging, prevalent methodologies include radiomics and deep learning. Previous studies have delved into using deep learning techniques to predict ICIs effectiveness and anticipate the expression of immune markers such as tumor mutation burden (TMB), programmed cell death 1 ligand 1 (PD-L1), and tumor mircroenvirment (TME) [13, 14].

Unlike earlier radiomics studies focusing on entire tumors, a burgeoning approach known as habitat imaging specifically targets identifying distinct tumor areas or cell subpopulations. Habitat radiomics analysis is the study of the living environment of organisms in ecological studies, while in radiomics, we can analogy it to the biological microenvironment around a tumor. Habitat radiomics analysis is a clustering of similar features based on traditional radiomics aimed at clearly identifying different tumor subregions or cell subpopulations [15, 16]. Some authors have detailed histologic validation of radiological habitat through per-pixel spatial co-registration of images with corresponding histologic findings of hypoxia, necrosis, and other conditions [17].

Traditional radiomics obscure the complexity of the immune microenvironment within the tumor, hindering a more in-depth study of the effect of intratumor heterogeneity (ITH) on therapeutic outcomes. Therefore, quantitative measurement of ITH may be a valuable biomarker for predicting clinical outcomes.

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 3 of 19

This study pioneers the use of multicenter data to apply habitat-based radiomics techniques, aiming to predict the emergence of ICIs effects and immune-related adverse events through pretreatment imaging studies.

Methods
Patients' cohort and images acquisition
Baseline chest CT images and clinical data of 248 patients were retrospectively collected from Jiangsu Provincial People's Hospital (the First Affiliated Hospital of Nanjing Medical University) between December 2018 and October 2023. All eligible patients underwent pretreatment chest CT scans 8 weeks before initiating immunotherapy. These images were then randomly divided into two cohorts: the training cohort (n = 173) and the validation cohort (n = 75). Furthermore, an external validation cohort was retrospectively assembled from 20 patients at Nanjing Pukou People's Hospital between December 2019 and July 2023. This multicenter retrospective study adhered to the Declaration of Helsinki and received approval from the Institutional Review Board of Jiangsu Provincial People's Hospital (SMC 2017-08-136). Given the retrospective nature of the study and the utilization of anonymized clinical data, informed consent was not required.

The inclusion criteria for this study were as follows: (1) histopathologically confirmed NSCLC diagnosis; (2) treatment with ICIs monotherapy (200 mg every 3 weeks at the approved dose) or ICIs combined with chemotherapy as first-line treatment; (3) clear documentation of endpoint events and status; (4) complete pretreatment CT scans; and (5) comprehensive clinical data, including gender, age, smoking status, TNM stage, pathological type, and tumor treatment regimen.

The exclusion criteria were: (1) incomplete image sequences; (2) a history of lung lesion surgery before or during immunotherapy follow-up; and (3) absence of measurable lesions in the pulmonary window as per RECIST 1.1 criteria. Figure 1 shows the inclusion and exclusion criteria for patients.

Clinical data, including age, sex, stage, histological type, progression-free survival (PFS), and irAEs, were extracted from medical records. Patient staging was conducted in accordance with the eighth edition of the American Joint Committee on Cancer criteria. Tumor response to immunotherapy post-treatment was evaluated using the Response Evaluation Criteria in Solid Tumours (RECIST version 1.1) criteria. Disease progression (PD) was defined as a 20% increase in target lesion volume or the emergence of a new lesion, while disease stabilization (SD) was also noted. The study's primary endpoint was PFS, measuring the duration from the initiation of immunotherapy to disease progression or death.

Patients underwent CT scans using various machines: SIEMENS SOMATOM Definition AS+, SIEMENS SOMATOM Definition Force, SIEMENS SOMATOM go.Up, SIEMENS Emotion 16, GE MEDICAL SYSTEMS Revolution CT, GE MEDICAL SYSTEMS Optima CT520 Series, and Philips iCT, with a section thickness of 1.5 or 5 mm for pulmonary window CT images. Clinical and demographic data were meticulously gathered from electronic medical records by two independent investigators, YQ Jiang and WX Zhou. Adverse events were assessed in accordance with the National Cancer Institute Common Terminology Criteria for Adverse Events, version 5.0. Any discrepancies were resolved through consultation with a third investigator, YM Wu. The study design and workflow are illustrated in Fig. 2.

Data preprocessing
In this study, we standardized the pixel values to a range of −1000 to 2000. It helps mitigate distorted analysis or inaccuracies caused by extreme pixel values. Maintaining appropriate pixel value ranges is crucial in medical imaging, as variations in contrast and intensity can significantly impact diagnostic precision.

To ensure uniform voxel spacing, we employed fixed resolution resampling to achieve voxel dimensions of $2\,mm \times 2\,mm \times 5\,mm$. Consistency in voxel sizes is paramount for maintaining comparability across different images. Standardizing these parameters facilitates more dependable comparisons and precise measurements.

Habitat based radiomics procedure
Habitat generation
Local features, including local entropy and energy values, were computed for each voxel within the Volume of Interest (VOI) in this study. These computations generated feature vectors that encapsulated a variety of voxel characteristics. Utilizing a $3 \times 3 \times 3$ moving window, these features were calculated for every voxel in the CT images, as illustrated in Fig. 3a, b, resulting in a 19-dimensional feature vector for each voxel.

Subsequently, the K-means clustering algorithm was employed to partition the VOI into three distinct regions for each sample, with the aim of preventing over-parameterization. This decision was influenced by prior research in habitat-related studies, which led to the identification of three clustering centers corresponding to the number of habitat regions. We propose that exploring different numbers of clustering centers in future experiments may enhance the model's performance.

Feature extraction
The handcrafted features in this study are classified into three categories: geometry, intensity, and texture.

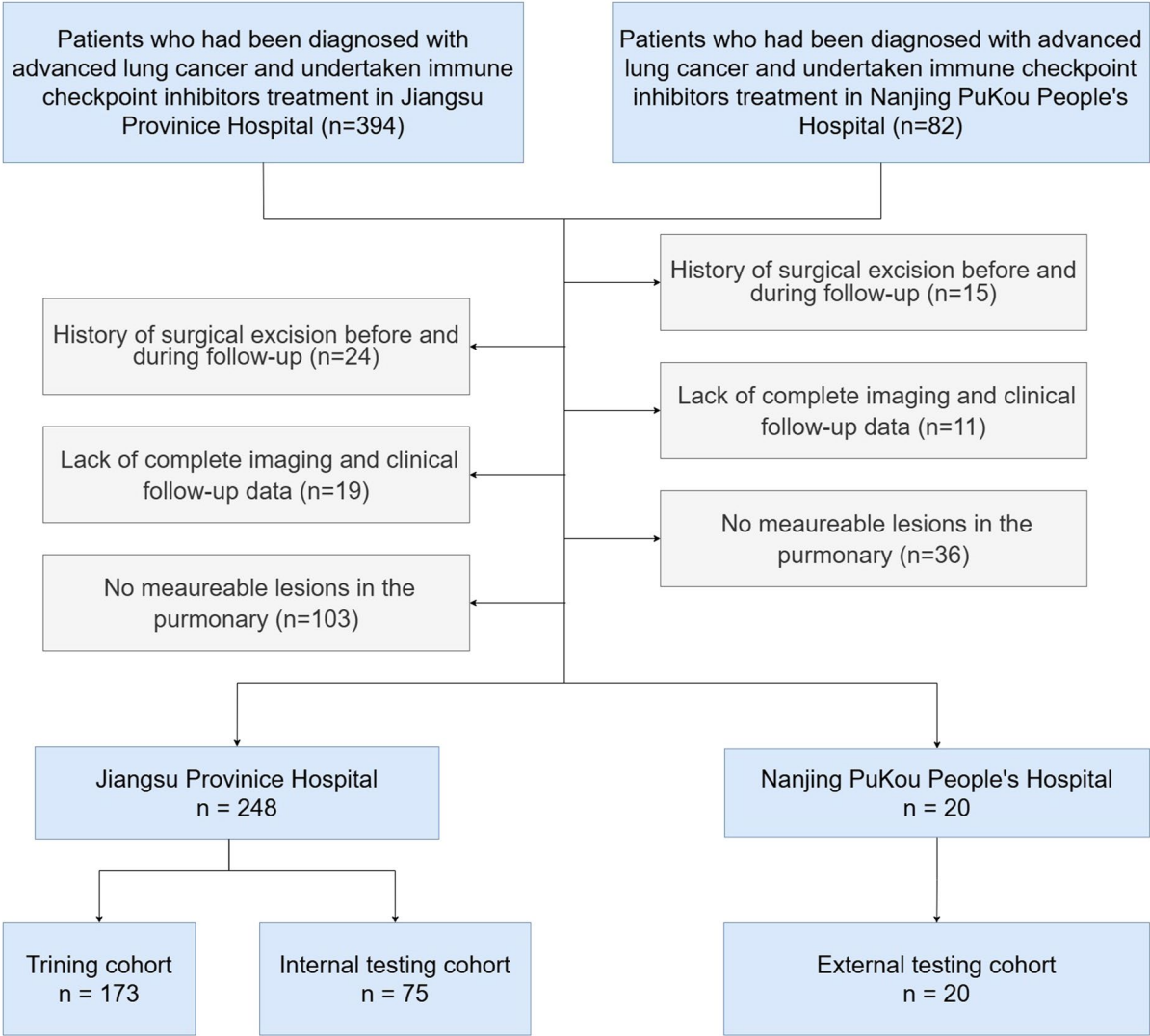Wu *et al. Journal of Translational Medicine* (2025) 23:393

Page 4 of 19

Fig. 1 The flowchart of inclusion criteria of patients

Geometry features describe the three-dimensional shape of the tumor, while intensity features offer a statistical analysis of voxel intensities. Texture features, on the other hand, aim to capture spatial patterns of intensity. Various methodologies, such as GLCM, GLRLM, GLSZM, and NGTDM, were employed to extract these texture features.

In the analysis of intra-tumor characteristics, where the Volume of Interest (VOI) was treated as a unified entity, the aforementioned radiomics features were extracted. Specific features were extracted for each identified subregion for habitat features.

Due to the unsupervised nature of the clustering algorithm used, ensuring consistent labeling of each subregion post-clustering was challenging. To address this issue, the ReMap technique was applied to remap similar areas, aiming to maintain consistent physical significance among different habitat regions, as illustrated in Supplementary 1B. All features were extracted using the pyradiomics tool (version 3.0.1), with most features conforming to the definitions outlined by the Imaging Biomarker Standardization Initiative (IBSI).

Feature selection

To achieve a normal distribution, all features underwent standardization using the mean and standard deviation derived from the training cohort, a process subsequently applied to the test cohort. Following this standardization, p-values for all imaging features were
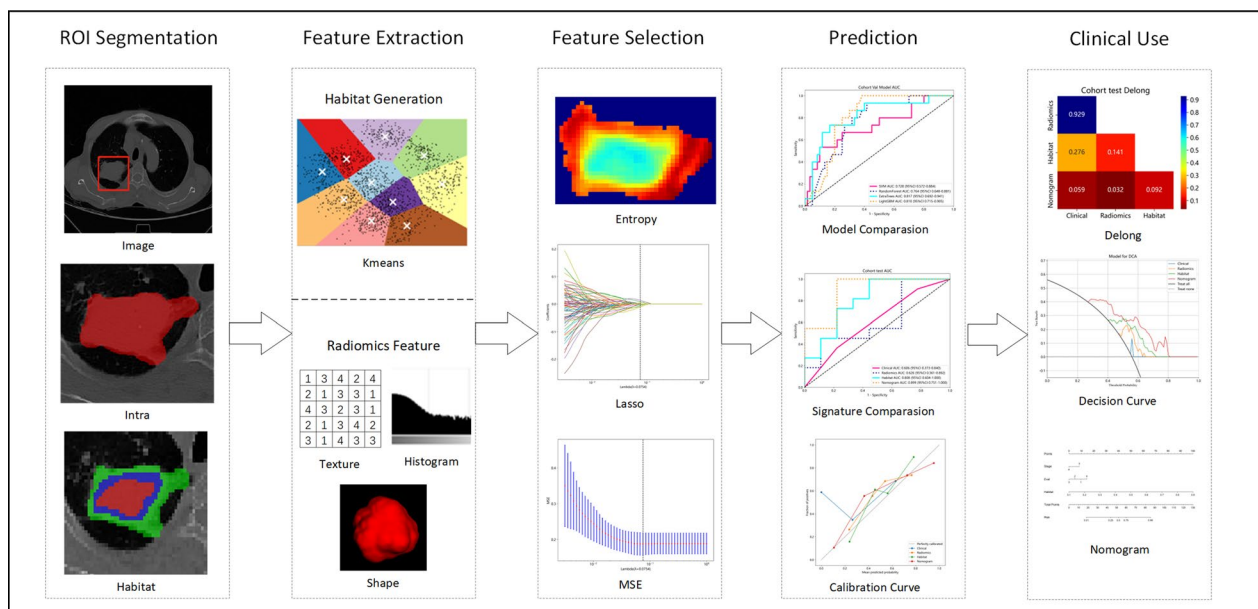
Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 5 of 19

Fig. 2 Overall workflow of this study

computed using a t-test, with only features having a *p*-value below 0.05 retained.

For features exhibiting high repeatability, further analysis was conducted using Pearson's correlation coefficient to identify strong correlations. In instances where the correlation coefficient between any two features exceeded 0.9, only one of the features was preserved. A recursive deletion approach was employed to enhance feature representation while minimizing redundancy, systematically eliminating the feature with the highest redundancy in each iteration.

The final selection of features for constructing the radiomic signature was determined through the LASSO (Least Absolute Shrinkage and Selection Operator) regression model. LASSO regression operates by shrinking regression coefficients towards zero, effectively discarding irrelevant features by setting their coefficients to zero based on the regularization weight λ. The optimal λ value was identified through tenfold cross-validation, selecting the value corresponding to the lowest mean standard error.

Radiomics signature

This study examined the effectiveness of different tumor region analyses in predicting Immunoreaction and Immune Efficacy. The analyses compared included assessing the tumor region as a whole (Whole) and analyzing tumor habitats (Habitat).

For the Whole-tumor Radiomics Signature, following Lasso feature screening, the selected features were utilized in machine learning methods to generate the radiomics signature. Widely accepted machine learning models, such as Logistic Regression (LR) for linear models and XGBoost for tree-based models, were employed to develop the risk model.

Regarding the Habitat Signature, internal tumor habitat characterization relied on unsupervised clustering algorithms, which could result in clusters with identical centers not necessarily holding the same physical meaning. To mitigate this, we calculated the meaning value of the features. The remaining steps in this process mirrored those used in the whole tumor radiomics models.

Model evaluation

We evaluated model efficacy for both adverse reactions and prognostic outcomes, with PFS segmented at seven months. We classified patients with PFS greater than 7 months as having a good prognosis, and patients with PFS less than or equal to 7 months as having a poor prognosis.

In the Clinical Signature analysis, we examined the significance of clinical features by incorporating all available parameters into our model. Due to the limited number of clinical features, we refrained from implementing a selection process at this stage. The methodologies applied here were consistent with those used in the radiomic and habitat modeling processes.

For the Radiomics Nomogram Signature, a nomogram was developed in this study that combines clinical features with handcrafted habitat-based radiomic
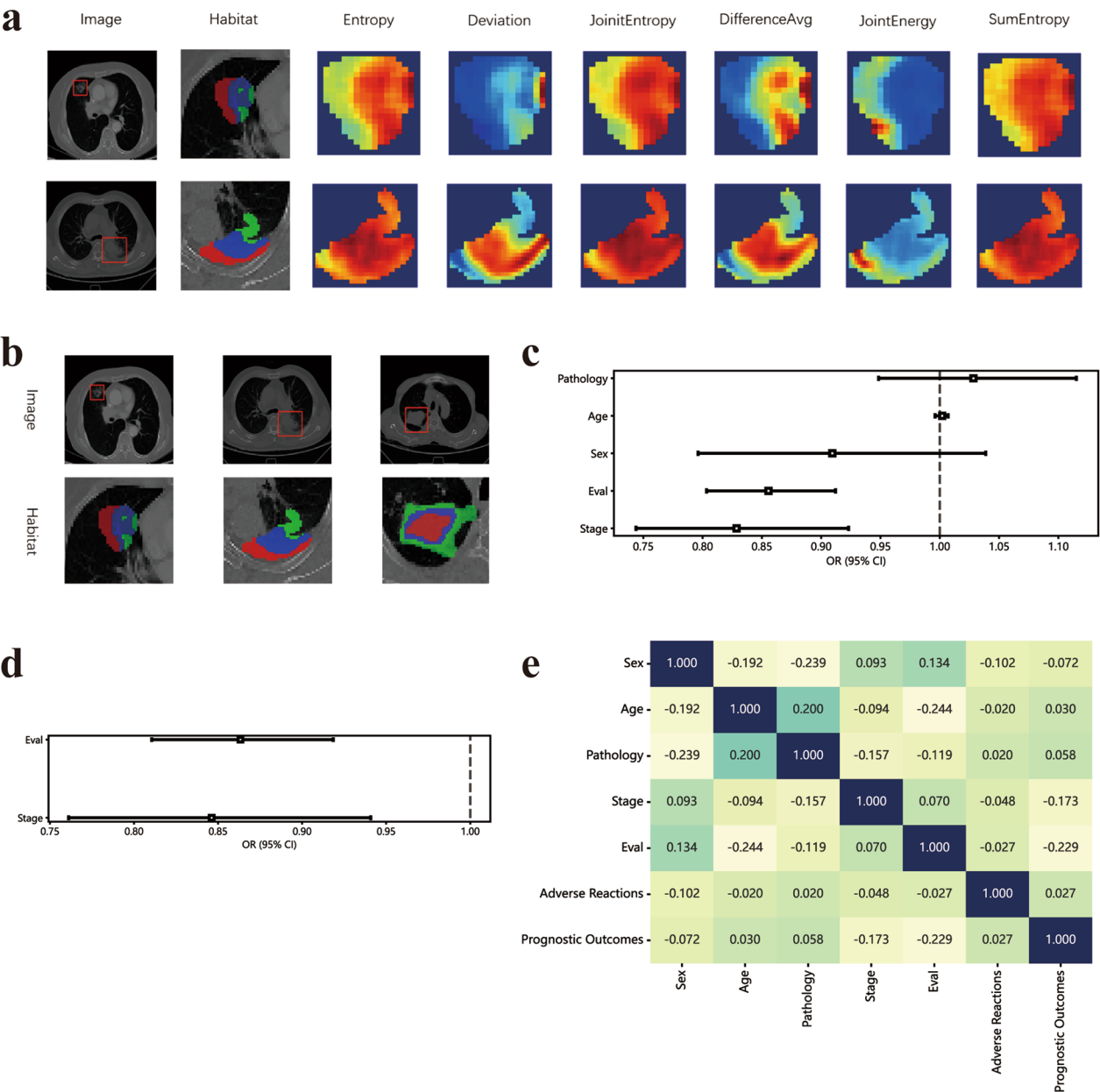
Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 6 of 19

Fig. 3 **a** presents the peritumoral regions generated and the habitat regions generated. **b** displays various samples with three cluster labels, yet the assigned labels differ. Our clustering algorithm categorized the regions into inner, middle, and edge regions. In the illustration, the red area on the left corresponds to the middle region, while the green area on the right corresponds to the middle region. **c** OR of clinical features in univariable analysis, **d** OR of clinical features in multivariable analysis. **e** Spearman correlation coefficients of each clinical features

signatures. The diagnostic performance of this nomogram was assessed in the test cohort using receiver operating characteristic (ROC) curves. Calibration curves were generated to evaluate calibration accuracy, and the Hosmer–Lemeshow goodness-of-fit test was employed to further assess calibration capability. Calibration curves assess the agreement between predicted probabilities and actual outcomes. They plot the predicted probability of an event against the observed frequency of that event. A well-calibrated model is represented by a curve that closely follows the 45-degree line, indicating that predicted probabilities align with observed outcomes. Calibration curves are essential for understanding how accurately a model predicts real-world events and are particularly important for ensuring the reliability of risk predictions in clinical settings. Additionally, DCA was

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 7 of 19

conducted to assess the clinical utility of the predictive models. DCA is a method used to evaluate the clinical utility of a predictive model by quantifying the net benefit across different probability thresholds. It compares the model's performance with simpler strategies, such as treating all patients or treating none, and helps determine the threshold at which the model provides the most value for decision-making. A model with a higher net benefit at a given threshold is considered more clinically useful.

Statistical analysis

Statistical tests were conducted to determine the normality of clinical features, utilizing the Shapiro method. For those features adhering to a normal distribution, the ttest was applied to perform significance analysis. In contrast, for features not following a normal distribution, the χ^2 test was used to compare the clinical characteristics of patients.

To pinpoint significant clinical features, both univariate and stepwise multiple regression analyses were employed for feature selection. These analyses were conducted using Python version 3.7.12 and statsmodels version 0.13.2. Furthermore, the development of machine learning models was carried out utilizing the scikit-learn version 1.0.2 interface.

Results

Patient cohort and clinical features

A total of 268 patients were retrospectively screened. Table 1 presents the clinicopathologic characteristics of the patients. The training cohort consisted of 173 patients (141 men and 32 women; mean age 63.40 ± 8.87 years), the validation set included 75 patients (63 men and 12 women; mean age 62.81 ± 9.68 years), and the external validation set comprised 20 patients (17 men and 3 women; mean age 68.70 ± 11.18 years). Among all patients, 144 had LUSC, 105 had LUAD, and 19 had other types of NSCLC pathology. There were 82 patients in stage 3 and 186 in stage 4. Univariate analysis was conducted on all clinical features, with the calculation of the Odds ratio (OR) and corresponding *p*-values for each feature. Notably, the feature Stage and Eval demonstrated *p*-values below 0.05 and was therefore selected for inclusion in the construction of the nomogram. The results below pertain to the univariate and multivariate analyses for the task of prognostic outcomes. No significant clinical features were identified in the analysis of adverse reactions, and hence, in the construction of the Nomogram, we focused solely on the analysis of prognostic outcomes (Table 2).

Patients using immune checkpoint inhibitors are evaluated for efficacy every 2–3 months. We define PR patients as a "good evelution", SD patients as a "medium evelution" and PD patients as a "bad evelution". In the univariate analysis, poor evaluation (HR 0.856, 95% CI 0.803–0.912; *P* < 0.05) and stage (HR 0.829, 95% CI 0.744–0.923; *P* < 0.05) were independently linked to shorter PFS. Subsequently, both evaluations (HR 0.846, 95% CI 0.761–0.941; *P* < 0.05) and stage (HR 0.863, 95% CI 0.811–0.919; *P* < 0.05) emerged as significant factors associated with PFS in the multivariable analysis (Fig. 3c–e).

Features selection

In the radiomics signature, a comprehensive set of 1834 handcrafted radiomic features was extracted and

Table 1 Baseline characteristics of the patients

Feature_name	ALL	Train	Val	Test
Age	63.63 ± 9.36	63.40 ± 8.87	62.81 ± 9.68	68.70 ± 11.18
Sex				
Male	221(82.46)	141(81.50)	63(84.00)	17(85.00)
Female	47(17.54)	32(18.50)	12(16.00)	3(15.00)
Pathology				
LUAD	144(53.73)	91(52.60)	46(61.33)	7(35.00)
LUSC	105(39.18)	73(42.20)	24(32.00)	8(40.00)
Other	19(7.09)	9(5.20)	5(6.67)	5(25.00)
Stage				
III	82(30.60)	52(30.06)	28(37.33)	2(10.00)
IV	186(69.40)	121(69.94)	47(62.67)	18(90.00)
Eval				
0	2(0.75)	2(1.16)	Null	Null
1	42(15.67)	26(15.03)	10(13.33)	6(30.00)
2	21(7.84)	8(4.62)	4(5.33)	9(45.00)
3	203(75.75)	137(79.19)	61(81.33)	5(25.00)

Table 2 Univariable and Multivariable Analysis of clinical features

Feature_name	OR	OR lower 95%CI	OR upper 95%CI	*p*_value	OR	OR lower 95%CI	OR upper 95%CI	*p*_value
Stage	0.829	0.744	0.923	0.004	0.846	0.761	0.941	0.01
Eval	0.856	0.803	0.912	0.000	0.863	0.811	0.919	0.0
Sex	0.910	0.796	1.039	0.239				
Age	1.002	0.996	1.007	0.645				
Pathology	1.029	0.948	1.115	0.564				

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 8 of 19

categorized into three primary groups: shape, first-order, and texture. This included 360 first-order features, 14 shape features, and a variety of texture features. For habitat-based radiomics signatures, we utilized the multiTaskLassoCV method, enabling multi-task learning within a single model to fit our data. In the clinical signatures, five clinical parameters—gender, age, stage, pathology, and evaluation—were incorporated into the modeling process.

This methodology, along with tenfold cross-validation, was employed for selecting both tumor and habitat-related features. A visual representation of this process is provided in Fig. 4.

Radiomics results

The radiomics signature models demonstrated strong performance in predicting irAEs, with the best AUC values being 0.976 (95% CI 0.948–1.000) for the training cohort, 0.609 (95% CI 0.436–0.782) for the validation cohort, and 0.720 (95% CI 0.426–1.000) for the external validation cohort. For prognostic outcome prediction, the radiomics signature model achieved optimal AUC values of 0.994 (95% CI 0.987–1.000) for the training cohort, 0.736 (95% CI 0.619–0.853) for the validation

cohort, and 0.687 (95% CI 0.436–0.938) for the external validation cohort (Fig. 5).

In comparison, the habitat signature model outperformed the radiomics signature model in predicting irAEs, with superior AUC values of 0.989 (95% CI 0.977–1.000) for the training cohort, 0.817 (95% CI 0.692–0.941) for the validation cohort, and 0.800 (95% CI 0.599–1.000) for the external validation cohort. Regarding prognostic outcome prediction, the habitat signature models achieved the best AUC values of 0.987 (95% CI 0.976–0.998) for the training cohort, 0.804 (95% CI 0.705–0.903) for the validation cohort, and 0.808 (95% CI 0.604–1.000) for the external validation cohort, surpassing the radiomics signature model (Fig. 6). However, the clinical model exhibited poor predictive performance, as depicted in Fig. 7.

The incorporation of tumor sub-region features and clinical characteristics into the nomogram signature significantly enhanced the prediction of immune-related adverse reactions, achieving AUC values of 0.923 (95% CI 0.883–0.963), 0.817 (95% CI 0.719–0.914), and 0.899 (95% CI 0.751–1.000) in training, validation, and external test cohorts, respectively. Simultaneously, the nomogram signature excelled in prognostic outcome prediction with
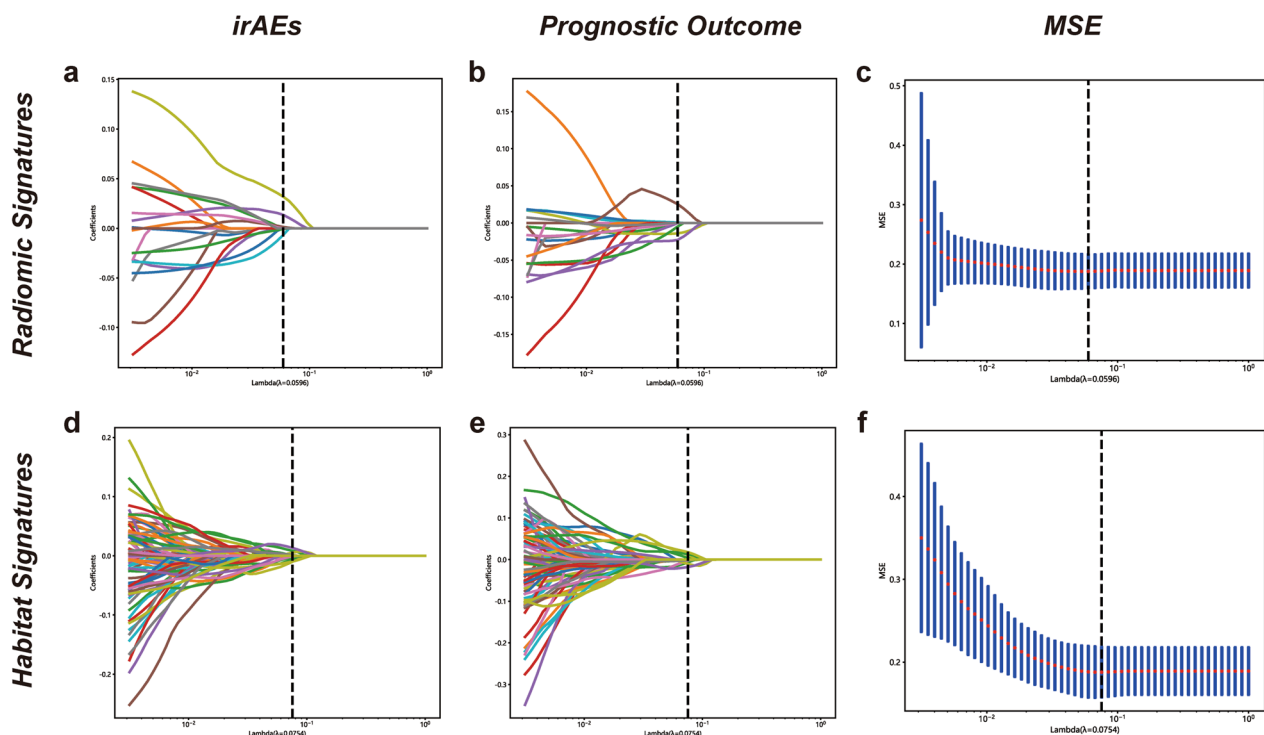

Fig. 4 **a**–**b**, **d**–**e** presents the coefficients obtained from LASSO (Least Absolute Shrinkage and Selection Operator) during tenfold cross-validation, applied in both radiomics signature and habitat signature analyses. The left and middle columns of these sub-figures correspond to the tasks of adverse reactions and prognostic outcomes, respectively. Meanwhile, the top and bottom sections depict the Intra and habitat radiomic methodologies, respectively. This layout effectively illustrates the differences and similarities in feature selection across the various tasks and methodologies. MSE path of lasso in tenfold cross validation in (**c**) radiomics signature and (**f**) habitat signature

Wu *et al. Journal of Translational Medicine*   (2025) 23:393
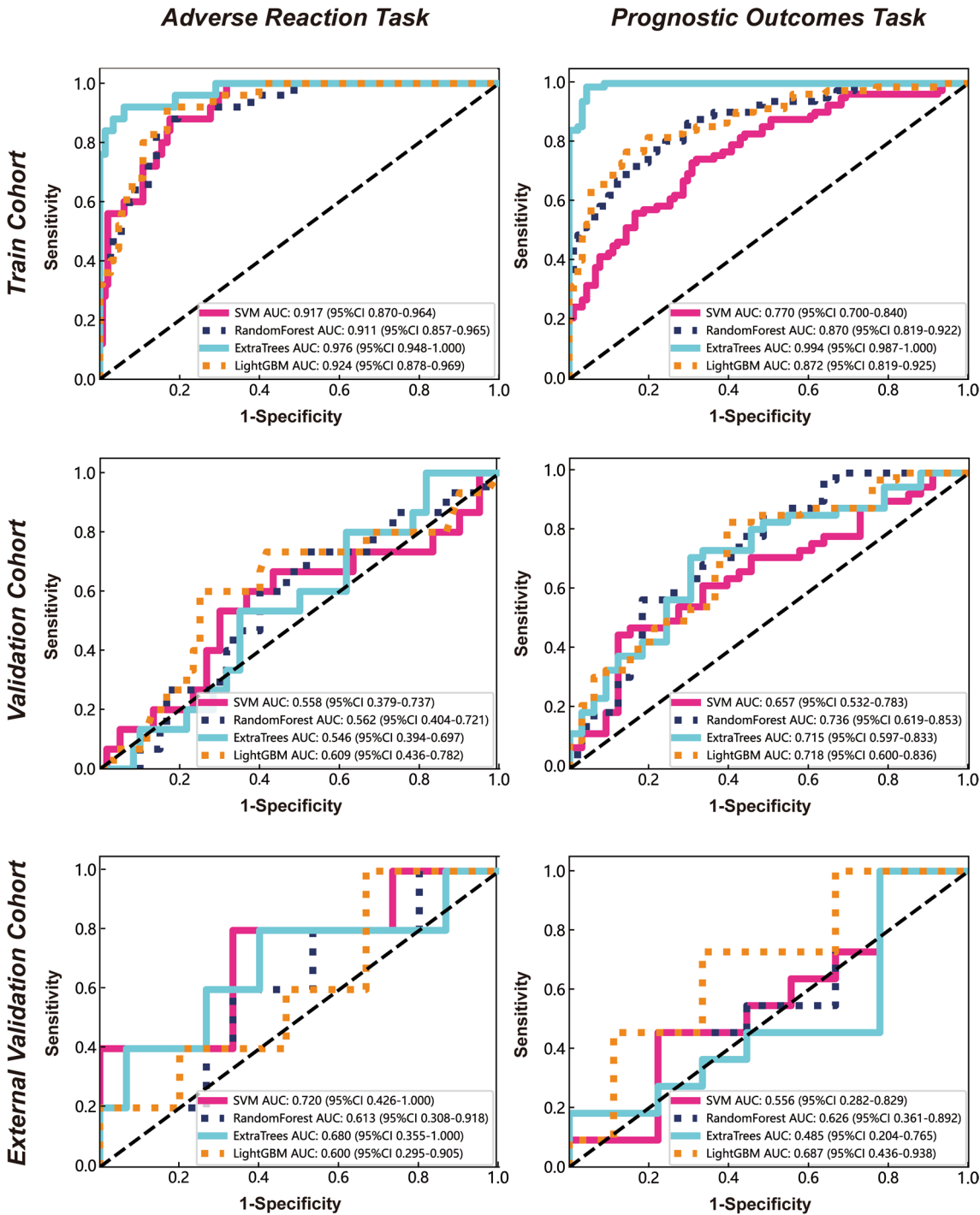
Page 9 of 19

Radiomic Signature

Fig. 5 Radiomic signature model performance analysis: The ROC curve for various models is presented along with the AUC and 95% confidence intervals shown in the bottom right. Left side: radiomic signature results for the adverse reaction task, right side: radiomic signature results for the prognostic outcome task
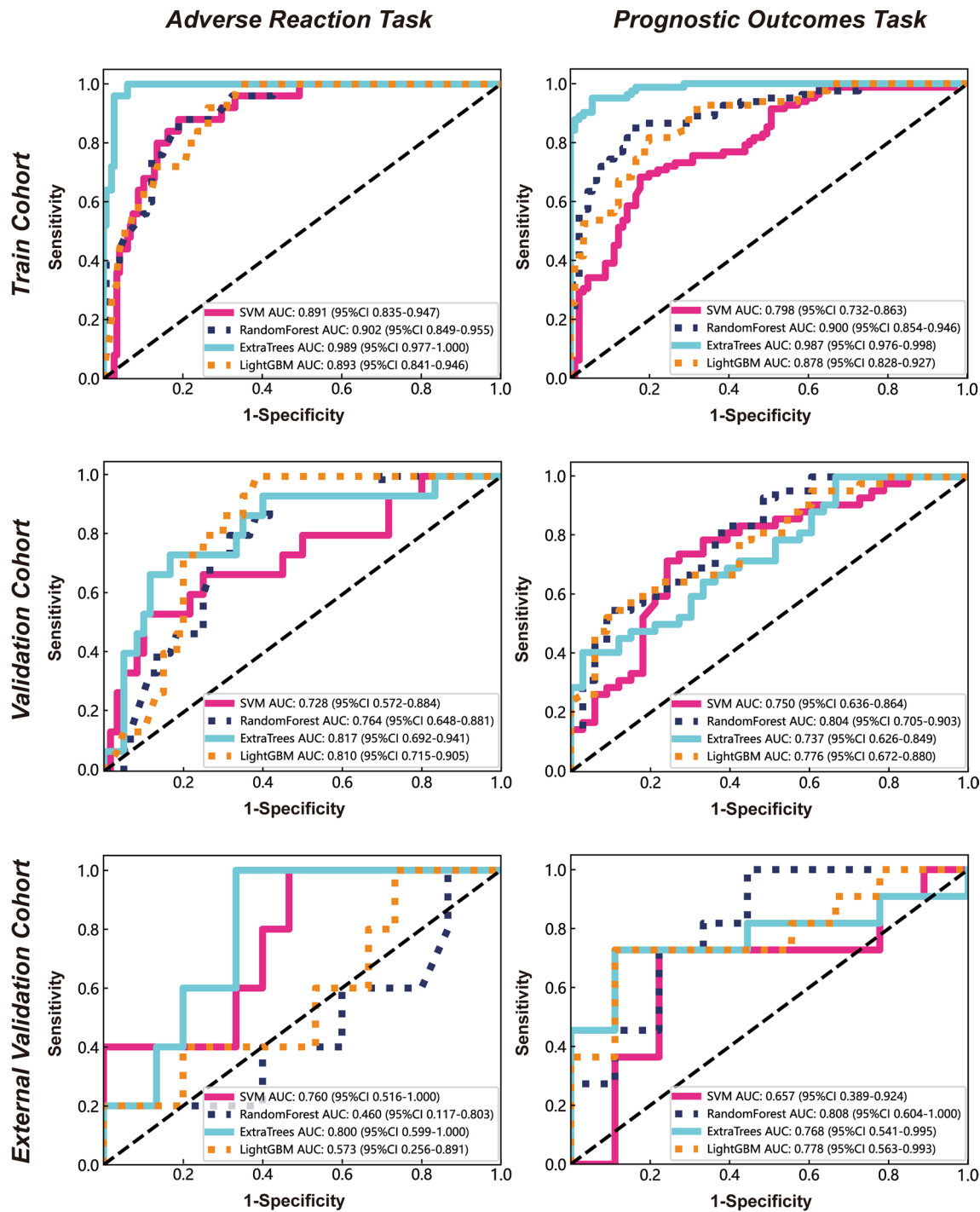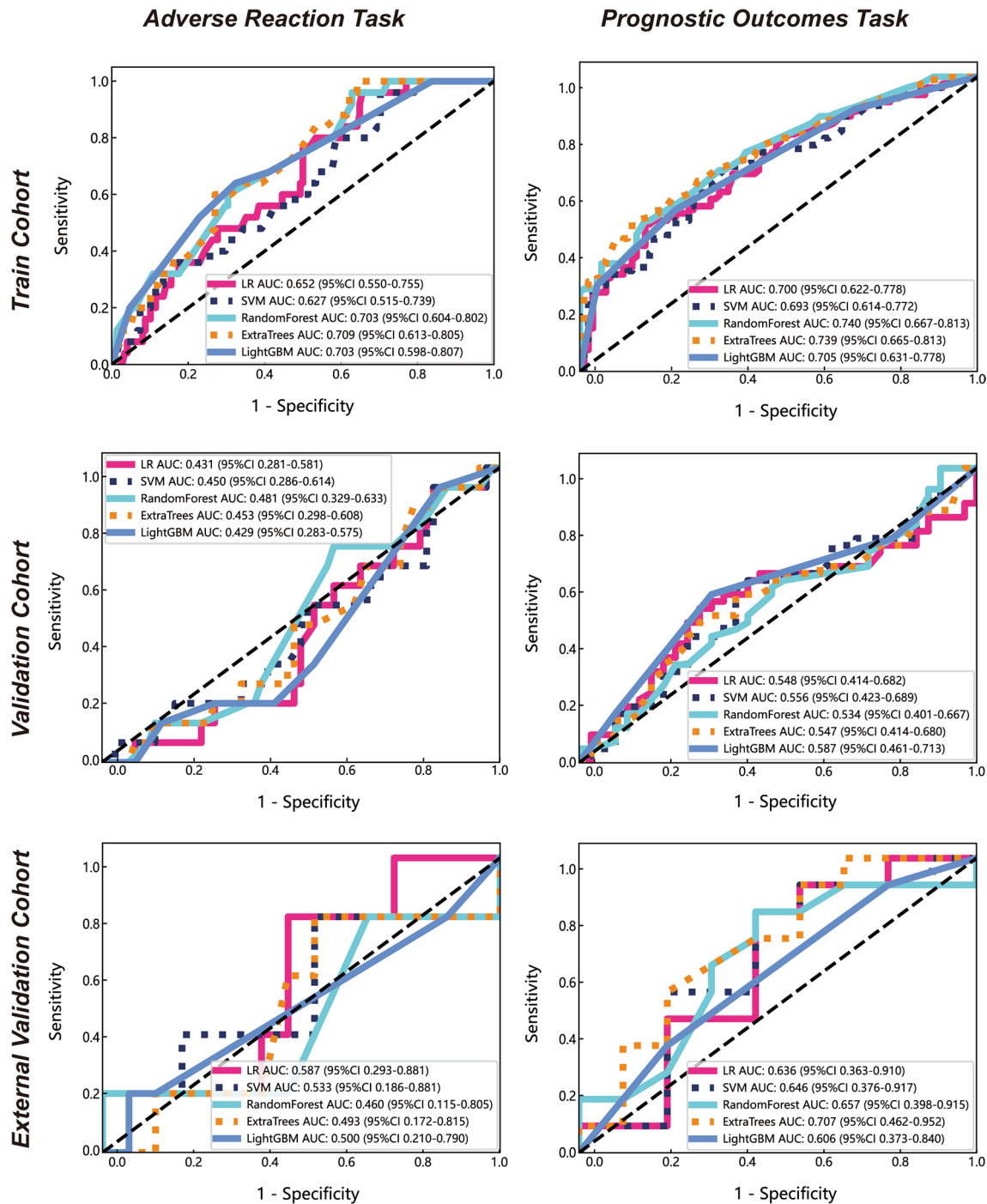
Habitat Signature

Fig. 6 Habitat radiomic signature model performance analysis: The ROC curve for various models is presented along with the AUC and 95% confidence intervals. Left side: habitat signature results for the adverse reaction task, right side: habitat signature results for the prognostic outcome task

Fig. 7 Clinical signature model performance analysis: The ROC curve for various models is presented along with the AUC and 95% confidence intervals. Left side: clinical signature results for the adverse reaction task, right side: clinical signature results for the prognostic outcome task

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 12 of 19

AUC values of 0.987 (95% CI 0.976–0.998), 0.804 (95% CI 0.705–0.903), and 0.808 (95% CI 0.604–1.000) in the training, validation, and external validation cohorts, respectively. The comparative results of different classifiers are illustrated in Fig. 8.

Model evaluation

Decision Curve Analysis (DCA) is a method used to evaluate the clinical utility of predictive models by considering the potential benefits and harms of applying a model in decision-making. Unlike traditional statistical measures (such as accuracy, sensitivity, or specificity), which only focus on model performance, DCA provides a framework for assessing the net benefit of a model at various decision thresholds. The DCA curve plots the net benefit of a model across a range of threshold probabilities, where the net benefit is calculated as the difference between the benefit of correctly identifying positive cases and the harm of false positives, weighted by the threshold probability of intervention. A model's net benefit is compared to two extreme strategies: always treating all patients (which maximizes true positives but also incurs high false positives) and never treating anyone (which avoids false positives but misses true positives). The threshold probability represents the point at which the benefits of a positive prediction outweigh the potential harm, and the decision curve provides insight into whether the model offers a clinically meaningful trade-off between these outcomes. A model with higher net benefit across a broader range of thresholds is considered more useful for guiding clinical decisions. In our study, decision curve analysis was performed to assess the clinical utility of these signatures by evaluating net benefits across various threshold probabilities. The analysis revealed that utilizing the radiomics Nomogram signature provided greater benefits in prediction compared to either the entire tumor radiomic signature or the habitat radiomics signatures (Fig. 9). The calibration curves of the nomogram model, as depicted in Fig. 10, underscore its robust clinical applicability. In the Delong test (Fig. 11), the diagnostic value of both habitat radiomic signatures and whole tumor radiomic signatures was significantly higher than that of clinical signatures, and the difference in AUC between the two was statistically significant ($p < 0.05$). Notably, both the habitat radiomics model and the radiomics model exhibited significantly higher diagnostic value than the clinical model, with notable differences in the AUCs of the models ($p < 0.05$). The Nomogram model demonstrated the highest diagnostic value among all models. It can be further refined as a quantitative tool for predicting the likelihood of irAEs occurrence and assessing immunotherapy efficacy

in NSCLC patients, presented in the form of a nomogram (Fig. 12).

Discussion

Significant advancements have been made in immunotherapy for malignant tumors; however, a substantial number of patients do not derive benefits from it. The intricate ITH poses a challenge to the comprehensive exploration of immunotherapy. Therefore, comprehending ITH is pivotal for assessing clinical efficacy and identifying novel targets for cancer immunotherapy [18, 19].

In contrast to prior radiomic investigations focusing on entire tumors, a novel approach has emerged to distinctly delineate various tumor regions or subpopulations of cells, often termed habitat analysis. There is a growing recognition of the dynamic interplay between tumors and their microenvironment (habitat)—where tumor cells can modify the microenvironment to promote their proliferation and survival. Concurrently, the habitat also influences the behavior of tumor cells [20]. Habitat radiomics distinguishes itself from traditional radiomics by focusing on the spatial heterogeneity within tumors. While conventional radiomics analyzes global tumor features such as shape, texture, and intensity, habitat radiomics segments tumors into distinct regions or "habitats," each reflecting varying biological characteristics, such as necrosis, hypoxia, and high cellularity. This approach recognizes that tumors are not homogeneous but consist of multiple microenvironments that influence treatment responses and outcomes. By considering the spatial distribution of features across these habitats, habitat radiomics offers a more detailed and dynamic understanding of tumor biology, which enhances predictive accuracy and enables more personalized treatment strategies [21, 22]. Utilizing multidimensional information and radiomic analysis for imaging habitat analysis is a common strategy. Studies have indicated that the relative habitat volume derived from clustering can serve as a prognostic indicator for survival. Some researchers have presented detailed preclinical findings and carried out histological validation of habitat analysis by meticulously aligning images pixel by pixel with corresponding histological observations [17, 23–25].

This study represents the initial endeavor, to our knowledge, to leverage intratumoral heterogeneity for exploring both irAEs and the efficacy of immunotherapy. This innovative approach stands to significantly aid patients in the selection and management of subsequent treatment regimens [26–30]. A predictive model utilizing features of intratumoral ecological diversity demonstrated strong performance in predicting the therapeutic response or irAEs to immunotherapy in NSCLC, validated internally and externally AUC. The optimal performance in
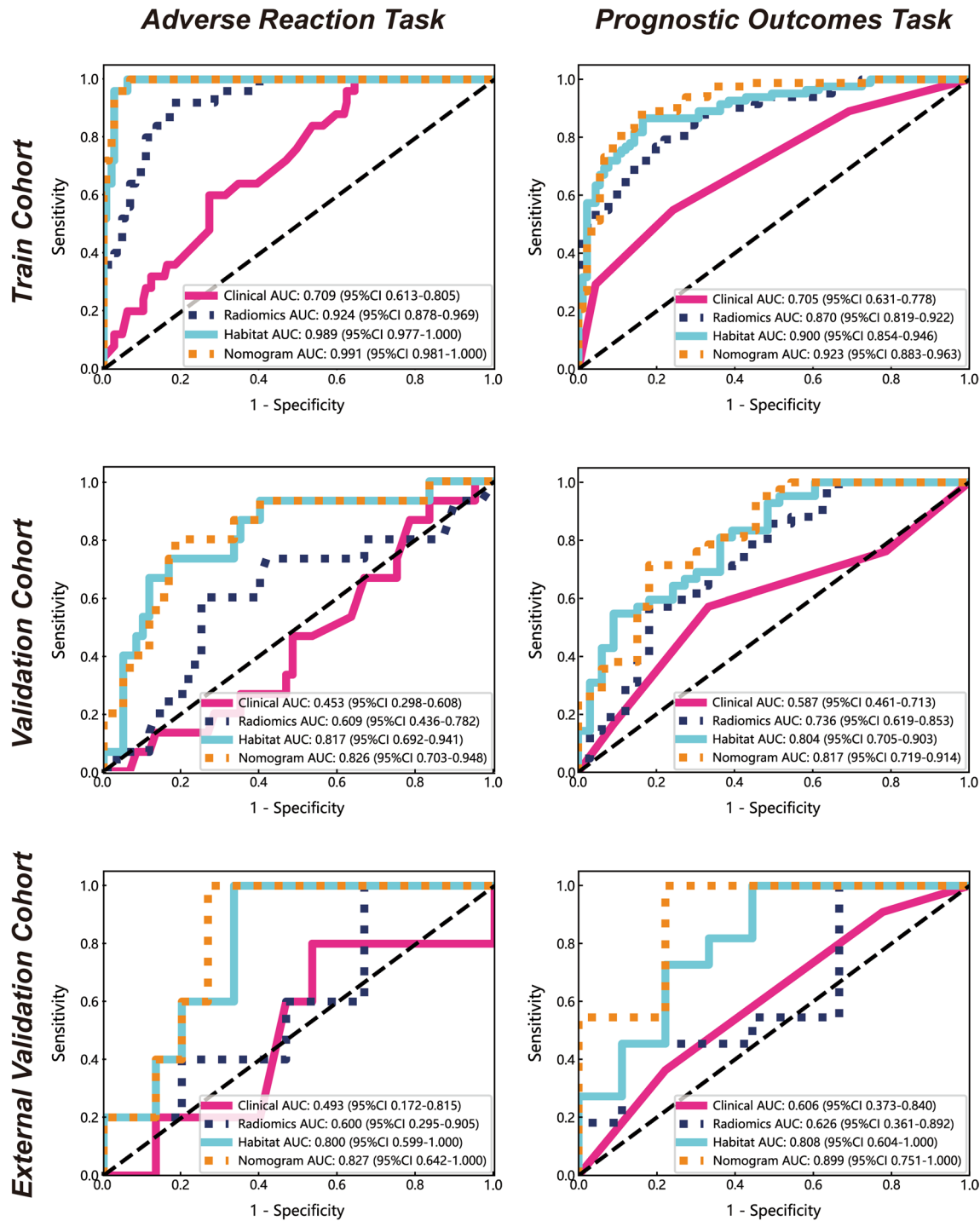
Fig. 8 Nomogram radiomic signature model performance analysis: The ROC curve for various models is presented along with the AUC and 95% confidence intervals. Left side: nomogram radiomic signature results for the adverse reaction task, right side: nomogram radiomic signature results for the prognostic outcome task
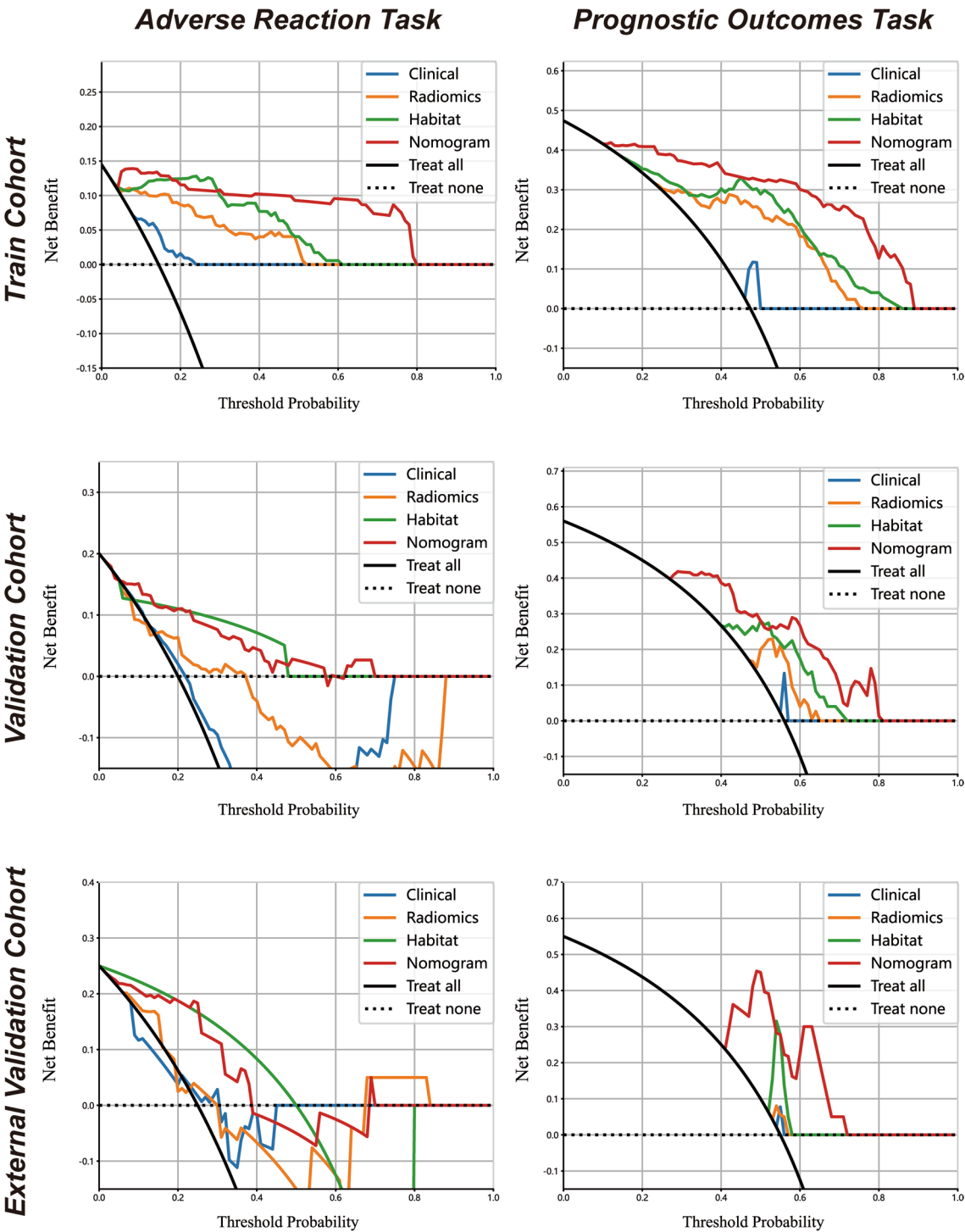
Decision Curve Analysis

Fig. 9 Figure 9 presents the decision curve analysis (DCA) for both the training and testing sets. The results indicate that our fusion model, the Nomogram, offers significant advantages in terms of predicted probabilities. Furthermore, when compared to other signatures, the nomogram displays greater potential for achieving net benefit, highlighting its effectiveness in clinical decision-making scenarios
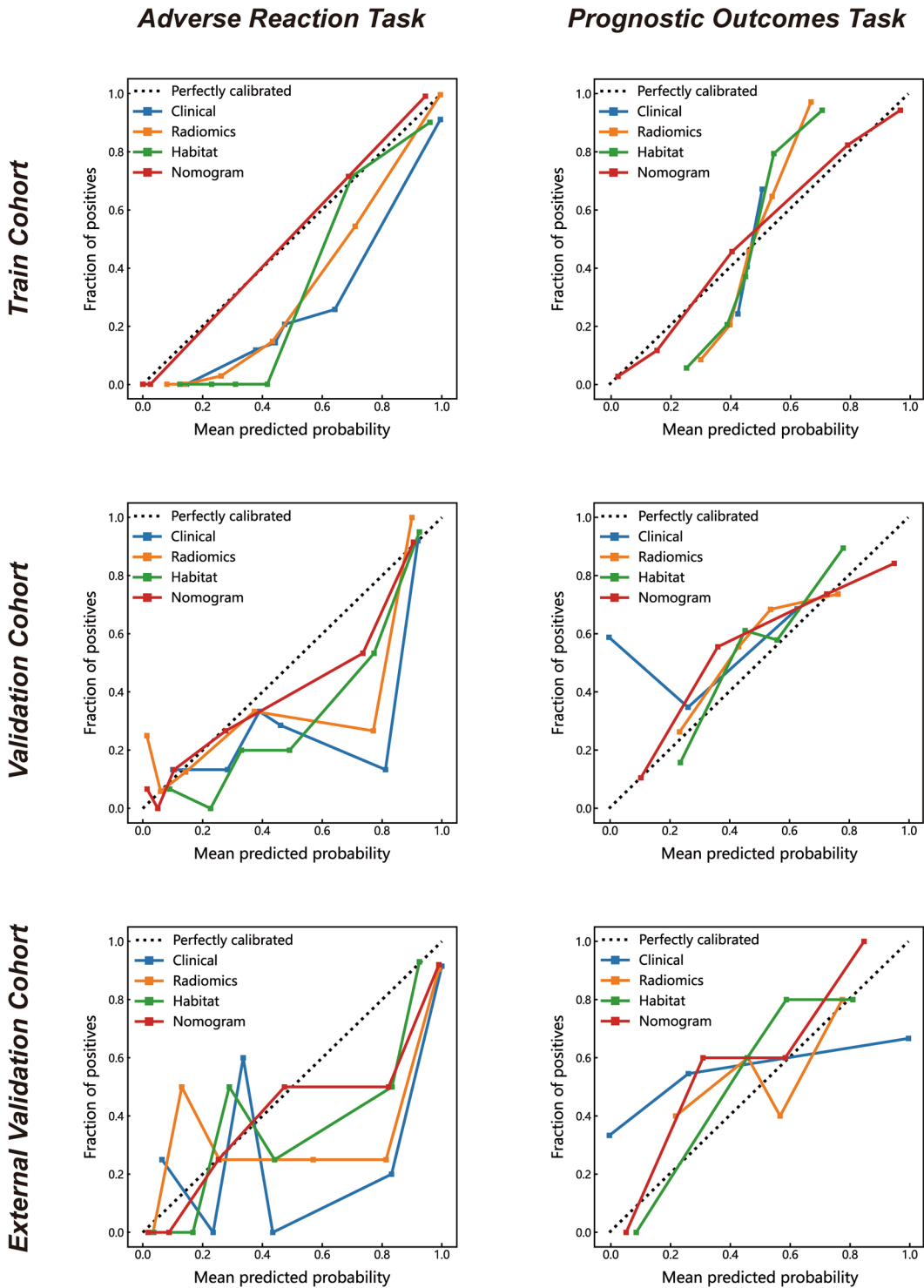
Calibration Curve

Fig. 10 The Hosmer–Lemeshow (HL) test statistics are used to measure the difference between the predicted probabilities of a model and the actual outcomes. A higher HL test statistic is indicative of better calibration, meaning the model's predictions more closely match the observed outcomes. In our study, the Nomogram model exhibited the highest calibration performance, as evidenced by its HL test statistics
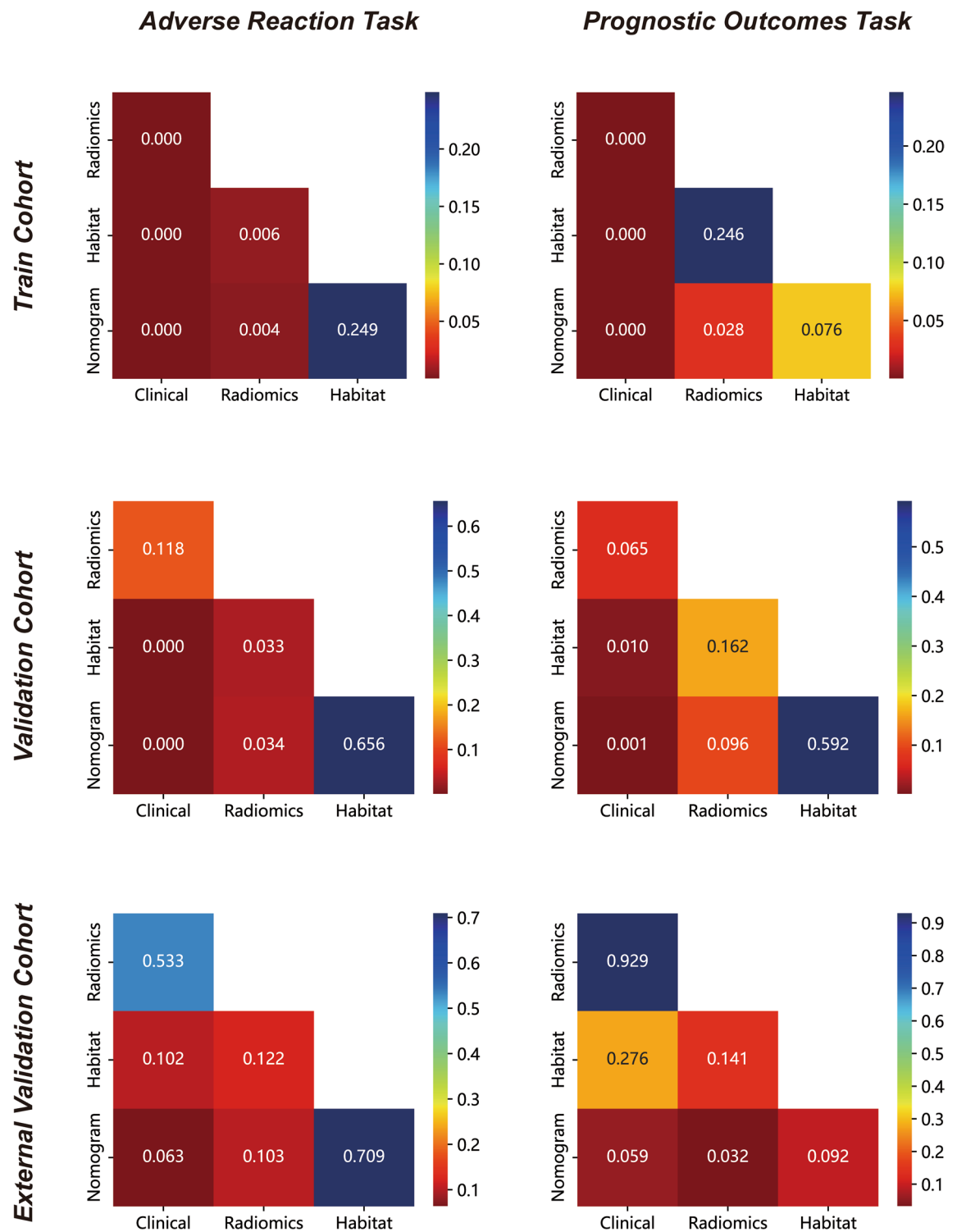
Delong Test

Fig. 11 Analysis of the training and testing datasets reveals that the Nomogram, which combines clinical features and habitat outcomes, outperforms other models. In the training cohort, all radiomic signatures significantly surpassed the Clinical Signature in terms of performance, as indicated by *p*-values less than 0.05. However, this marked improvement is not as evident in the test set. This discrepancy may be due to the relatively minor influence of the clinical model in the linear fusion process
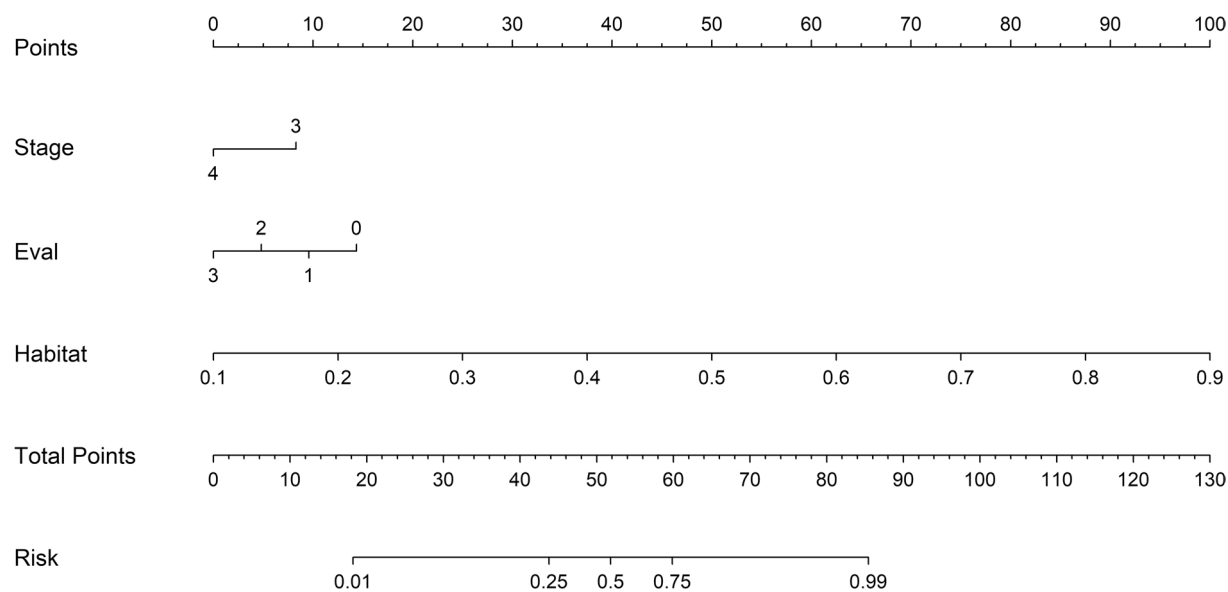
Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 17 of 19

Fig. 12 The nomogram for clinical use

forecasting the therapeutic response or irAEs to immunotherapy in external validation sets was achieved by integrating intratumoral ecological diversity features with clinical characteristics. The imaging signature not only offered independent prognostic insights surpassing established clinical factors but also stratified NSCLC patients into distinct risk categories. For patients with a low risk of irAEs and favorable immunotherapy efficacy, immunotherapy is advisable; conversely, for those with a high risk of irAEs and poor immunotherapy efficacy, ICIs should be cautiously administered with closer irAE monitoring. The imaging signature holds promise for refining patient selection among advanced NSCLC cases, distinguishing those who stand to benefit most from immunotherapy and those at heightened risk of adverse immune reactions.

Our study stands out in several key aspects. Firstly, we employed the innovative Habitat Fusion Architecture, a novel algorithmic framework that amalgamates clinical features with a comprehensive analysis of the tumor microenvironment. This method offers a detailed insight into the tumor habitat and demonstrates potential for accurately predicting immunological responses and the efficacy of immune therapies. Secondly, we utilized habitat techniques to integrate data from diverse tumor subregions, thereby enhancing the predictive capacity of our model. Thirdly, we developed a user-friendly nomogram that combines selected clinical features and tumor habitat indicators. This tool aims to improve prediction accuracy and assist in clinical decision-making, enhancing the accessibility and utility of the technology.

However, our study had several limitations. Firstly, retrospective radiomics research is prone to biases that can affect results. To minimize these, multi-center data collection improves representativeness, while standardizing preprocessing and expert annotation validation ensures consistency. Controlling confounders like age and disease stage helps avoid misleading conclusions. For robust model evaluation, cross-validation and external validation with independent datasets are essential. Addressing data imbalance through oversampling or weighted loss functions improves performance. These strategies reduce bias and enhance the clinical relevance of radiomics findings. Secondly, manual tumor delineation by different physicians could impact the stability of radiological features. Future studies should consider adopting automated tumor segmentation methods to enhance feature extraction efficiency and ensure process stability and consistency. Thirdly, notable differences existed in patient characteristics between the two hospitals, likely influenced by the distinct nature of each institution. Nevertheless, our hypothesis was statistically significant despite the heterogeneous patient cohort. Lastly, the complexity of the technology used in our study may render the technique impractical. Developing user-friendly software is essential to enhance clinical relevance. Nonetheless, our study serves as a proof-of-concept. Similar to advancements in artificial intelligence, accumulating large-scale data using various machines or scanning parameters will render this technique more practical.

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 18 of 19

Conclusion

Our study introduces a CT image-based Habitat technology machine learning model that effectively predicts irAEs and immunotherapy outcomes in patients with NSCLC. This model enables the identification of NSCLC patients at risk of developing irAEs or exhibiting positive responses to immunotherapy. These findings will undergo prospective validation in forthcoming multicenter, randomized trials to assess the clinical utility of our imaging model in guiding personalized treatment. By utilizing machine learning to forecast immunotherapy outcomes, our study advances the pursuit of precision immunotherapy for lung cancer, offering significant assistance to clinicians and patients in selecting and managing subsequent treatment regimens towards achieving personalized therapy goals.

Abbreviations

NSCLC	Non-small cell lung cancer
irAEs	Immune-related adverse reactions
ITH	Intratumor heterogeneity
ICIs	Immune checkpoint inhibitors
AUC	Area under operating characteristic curve
DCA	Decision Curve Analysis
IO	Immunotherapy
TMB	Tumor mutation burden
PD-L1	Programmed cell death 1 ligand 1
TME	Tumor mircroenvirment
PFS	Progression-free survival
RECIST version 1.1	Response Evaluation Criteria in Solid Tumours
PD	Disease progression
SD	Disease stabilization
VOI	Volume of Interest
IBSI	Imaging Biomarker Standardization Initiative
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Logistic Regression

Availability of data and materials

The code data presented in the study are openly available at https://github.com/SherryWym/Habitat-radiomics. The CT image datasets generated and/or analysed during the current study are not publicly available due to protect patient privacy but are available from the corresponding author on reasonable request.

Declarations

Author details

[1]Department of Oncology, First Affiliated Hospital, Nanjing Medical University, Nanjing, China. [2]Department of Radiology, First Affiliated Hospital, Nanjing Medical University, Nanjing, China. [3]Department of Lung Cancer Surgery, Tianjin Lung Cancer Institute, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China. [4]Department of Oncology, Pukou Branch of Jiangsu People's Hospital, Nanjing Pukou District Central Hospital, Nanjing, China. [5]Department of Pathology, First Affiliated Hospital, Nanjing Medical University, Nanjing, China.

References

1. Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. Cell. 2020;182(1).
2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018;553(7689):446–54.
3. Shankar B, Zhang J, Naqash AR, Forde PM, Feliciano JL, Marrone KA, et al. Multisystem immune-related adverse events associated with immune checkpoint inhibitors for treatment of non-small cell lung cancer. JAMA Oncol. 2020;6(12):1952–6.
4. Osmani L, Askin F, Gabrielson E, Li QK. Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): moving from targeted therapy to immunotherapy. Semin Cancer Biol. 2018;52(Pt 1):103–9.
5. Kennedy LB, Salama AKS. A review of cancer immunotherapy toxicity. CA Cancer J Clin. 2020;70(2):86–104.
6. Xu C, Chen Y-P, Du X-J, Liu J-Q, Huang C-L, Chen L, et al. Comparative safety of immune checkpoint inhibitors in cancer: systematic review and network meta-analysis. BMJ. 2018;363:k4226.
7. Novello S, Kowalski DM, Luft A, Gümüş M, Vicente D, Mazières J, et al. Pembrolizumab plus chemotherapy in squamous non-small-cell lung cancer: 5-year update of the phase III KEYNOTE-407 study. J Clin Oncol. 2023;41(11):1999–2006.
8. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer. 2017;16(1):41.
9. Song X, Xiong A, Wu F, Li X, Wang J, Jiang T, et al. Spatial multi-omics revealed the impact of tumor ecosystem heterogeneity on immunotherapy efficacy in patients with advanced non-small cell lung cancer treated with bispecific antibody. J Immunother Cancer. 2023;11(2):e006234.
10. Zhao X, Liang Y-J, Zhang X, Wen D-X, Fan W, Tang L-Q, et al. Deep learning signatures reveal multiscale intratumor heterogeneity associated with biological functions and survival in recurrent nasopharyngeal carcinoma. Eur J Nucl Med Mol Imaging. 2022;49(8):2972–82.
11. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal. 2022;79:102470.
12. Chen X, Wang X, Zhang K, Fung K-M, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79:102444.
13. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response

Wu *et al. Journal of Translational Medicine*   (2025) 23:393

Page 19 of 19

to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol. 2018;19(9):1180–91.

14. Park C, Jeong DY, Choi Y, Oh YJ, Kim J, Ryu J, et al. Tumor-infiltrating lymphocyte enrichment predicted by CT radiomics analysis is associated with clinical outcomes of non-small cell lung cancer patients receiving immune checkpoint inhibitors. Front Immunol. 2022;13:1038089.

15. Zhang X, Su G-H, Chen Y, Gu Y-J, You C. Decoding intratumoral heterogeneity: clinical potential of habitat imaging based on radiomics. Radiology. 2023;309(3):e232047.

16. Prior O, Macarro C, Navarro V, Monreal C, Ligero M, Garcia-Ruiz A, et al. Identification of precise 3D CT radiomics for habitat computation by machine learning in cancer. Radiol Artif Intell. 2024;6(2):e230118.

17. Park JE, Kim HS, Kim N, Park SY, Kim Y-H, Kim JH. Spatiotemporal heterogeneity in multiparametric physiologic MRI Is associated with patient outcomes in IDH-wildtype glioblastoma. Clin Cancer Res. 2021;27(1):237–45.

18. Hinohara K, Polyak K. Intratumoral heterogeneity: more than just mutations. Trends Cell Biol. 2019;29(7):569–79.

19. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81–94.

20. Tabassum M, Suman AA, Suero Molina E, Pan E, Di leva A, Liu S. Radiomics and machine learning in brain tumors and their habitat: a systematic review. Cancers. 2023;15(15):3845.

21. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. Clin Radiol. 2017;72(1):3–10.

22. Mu W, Liang Y, Hall LO, Tan Y, Balagurunathan Y, Wenham R, et al. 18F-FDG PET/CT habitat radiomics predicts outcome of patients with cervical cancer treated with chemoradiotherapy. Radiol Artif Intell. 2020;2(6):e190218.

23. Wu J, Cao G, Sun X, Lee J, Rubin DL, Napel S, et al. Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. Radiology. 2018;288(1):26–35.

24. Kim J, Ryu S-Y, Lee S-H, Lee HY, Park H. Clustering approach to identify intratumour heterogeneity combining FDG PET and diffusion-weighted MRI in lung adenocarcinoma. Eur Radiol. 2019;29(1):468–75.

25. O'Connor JPB, Rose CJ, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21(2):249–57.

26. Mu W, Jiang L, Zhang J, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. Nat Commun. 2020;11(1):5228.

27. Wang T, She Y, Yang Y, Liu X, Chen S, Zhong Y, et al. Radiomics for survival risk stratification of clinical and pathologic stage IA pure-solid non-small cell lung cancer. Radiology. 2022;302(2):425–34.

28. Le NQK, Kha QH, Nguyen VH, Chen Y-C, Cheng S-J, Chen C-Y. Machine learning-based radiomics signatures for EGFR and KRAS mutations prediction in non-small-cell lung cancer. Int J Mol Sci. 2021;22(17):9254.

29. Mu W, Katsoulakis E, Whelan CJ, Gage KL, Schabath MB, Gillies RJ. Radiomics predicts risk of cachexia in advanced NSCLC patients treated with immune checkpoint inhibitors. Br J Cancer. 2021;125(2):229–39.

30. Chen M, Lu H, Copley SJ, Han Y, Logan A, Viola P, et al. A Novel Radiogenomics biomarker for predicting treatment response and pneumotoxicity from programmed cell death protein or ligand-1 inhibition immunotherapy in NSCLC. J Thorac Oncol. 2023;18(6):718–30.

Publisher's Note