BMC Medical Imaging

RESEARCH

Open Access

Computed tomography radiomics reveals prognostic value of immunophenotyping in laryngeal squamous cell carcinoma: a comparison of whole tumor- versus habitats-based approaches

Meng Qi[1†], Weiding Zhou[2†], Ying Yuan[2], Yang Song[3], Duo Zhang[4*] and Jiliang Ren[2*]

Abstract

Background To compare the performance of whole tumor and habitats-based computed tomography (CT) radiomics for predicting immunophenotyping in laryngeal squamous cell carcinomas (LSCC) and further evaluate the stratified effect of the radiomics model on disease-free survival (DFS) and overall survival (OS) of LSCC patients.

Methods In all, 106 LSCC patients (40 with inflamed and 66 with non-inflamed immunophenotyping) were randomly assigned into a training ($n=53$) and testing ($n=53$) cohort. Briefly, 750 radiomics features from contrast-enhanced CT images were respectively extracted from the whole tumor and two Otsu method-derived subregions. Intraclass correlation coefficients (ICCs) were calculated to evaluate the reproducibility. The radiomics models for predicting immunophenotyping were respectively created using K-nearest neighbors (KNN), logistic regression (LR), and Naive bayes (NB) classifiers. The performance of models in the testing cohort were compared using area under the curve (AUC). The prognostic value of the optimal model was determined by survival analysis.

Results The radiomics features derived from whole tumor showed better reproducibility than those derived from habitats. The best model for the whole tumor (LR classifier) showed superior performance than that for the habitats (KNN classifier) in the testing cohort, but there were no significant differences (AUC: 0.741 vs. 0.611, $p=0.112$). Multivariable Cox regression analysis showed that the immunophenotyping predicted by the optimal model was an independent risk factor of unfavorable DFS ($p=0.009$) and OS ($p=0.008$) in LSCC patients.

Conclusions Whole tumor-based CT radiomics could serve as a potential predictive biomarker of immunophenotyping and outcome prediction in LSCC patients.

[†]Meng Qi and Weiding Zhou have contributed equally to this work.

*Correspondence:
Duo Zhang
doctorzhangduo@163.com
Jiliang Ren
renjiliang2016@163.com

Full list of author information is available at the end of the article

Keywords Larynx, Squamous cell carcinoma, Immunophenotyping, Radiomics, Computed tomography

Introduction

Head and neck cancers rank as the eighth-most fatal cancer globally, with 95% of these cases being squamous cell carcinomas [1]. Specifically, laryngeal squamous cell carcinomas (LSCC) are the second-most prevalent cancer of the head and neck [2]. Despite the application of improved multimodal therapies in LSCC over the past decade, the survival outcomes remain unsatisfactory [3, 4]. The immune microenvironment is widely recognized as a crucial factor in cancer progression, metastasis, and therapeutic response [5–8]. The immunoscore, a metric that summarizes the location and density of CD3+and CD8+T-cells, has gained widespread acceptance as an index for categorizing several cancers into inflamed and non-inflamed type [9, 10]. For LSCC, patients with the inflamed type exhibit longer survival and a lower recurrence rate compared to those with the non-inflamed type [11, 12]. In addition, immunophenotyping is associated with prediction of response to immunotherapy [13]. However, the immunoscore for LSCC is primarily determined through postoperative histological analysis, necessitating the development of noninvasive pretreatment tools.

Computed tomography (CT) has been used as a routine diagnostic tool for LSCC given its cost-effectiveness, rapid imaging capabilities, and minimal swallowing artifacts. Radiomics holds promise in extracting high-dimensional information from medical images, offering non-invasive means to capture tumor heterogeneity [14]. Previous researches have demonstrated the success of CT radiomics in predicting lymph node metastasis, gene status, and prognosis in LSCC patients [15–17]. Moreover, selected studies have hinted at the potential of radiomics in evaluating the immunophenotyping in gastric cancer [18], breast cancer [19], and hepatocellular cancer [20]. However, a common approach in radiomics research is to analyze the whole tumor, which may overlook local phenotypic variations within the tumor [21]. To address this, recent advancements have extended radiomics to evaluate subregions within the tumor, known as habitats. These habitats consist of voxels that share similar tumor biology and offer a deeper understanding of tumor heterogeneity. However, this approach requires more intricate image analysis techniques that may affect the reproducibility and practicability of modeling [21].

Therefore, we conducted a comparative analysis to assess the performance of CT radiomics models based on whole tumor and habitats for predicting immunophenotyping in LSCC. We also evaluated the prognostic significance of the predicted immunophenotyping to validate the potential clinical utility of radiomics models.

Materials and methods

Patients

This retrospective study received approval from the Ethics Review Board of Shanghai Eye & ENT Hospital, Fudan University; the requirement for informed consent was waived. The study included patients who underwent total or partial laryngectomy at the Otolaryngology-Head and Neck Surgery Department of Shanghai Eye & ENT Hospital between December 2014 and December 2022. The inclusion criteria specified that patients must have: (1) histopathological confirmation of LSCC following surgery; (2) undergone contrast-enhanced CT scanning prior to treatment; (3) at least one block of formalin-fixed and paraffin-embedded tissue available for tissue microarray; and (4) complete clinicopathological data. Patients were excluded if they exhibited: (1) images of poor quality due to artifacts or body motion; (2) tumors with a short axis less than 10 mm; or (3) recurrent tumor lesions. A total of 106 patients were enrolled in the study and randomly assigned to either a training cohort or a testing cohort in a 1:1 ratio. Survival information of 84 patients was obtained. Overall survival (OS) was defined as the start of treatment until the date of death, and disease-free survival (DFS) was calculated from the time from treatment initiation to tumor recurrence (local or distant) or death from non-tumor-related causes.

Immunohistochemical evaluation

The protocol for immunohistochemical staining of tissue microarray has been previously reported [11]. The stained tissue microarray sections were imaged using an Aperio digital slide scanner (Leica Biosystems) and subsequently analyzed with QuPath software (version 0.2.3). The quantification of tumor-infiltrating lymphocytes (TILs) involved summing the CD3+and CD8+T-cells present in the tumor interior (TI) and tumor stroma (TS). Patients were stratified into low or high-infiltrating groups based on the median number of TILs expressing CD3+and CD8+markers. The immunoscore for each patient was obtained by taking into account four infiltration levels (two markers and two regions). The classification of immunoscores was as follows: 0–2 indicated non-inflamed type, while 3–4 indicated inflamed type. The evaluation of CD3+and CD8+T-cells density and the calculation of immunoscore are shown in Fig. 1.

Image acquisition

Patients were scanned using 64-detector row CT machines (SOMATOM; Siemens Medical, Erlangen, Germany). The parameters for contrast-enhanced CT scanning were set at 120 kV, 200 mA, slice thickness

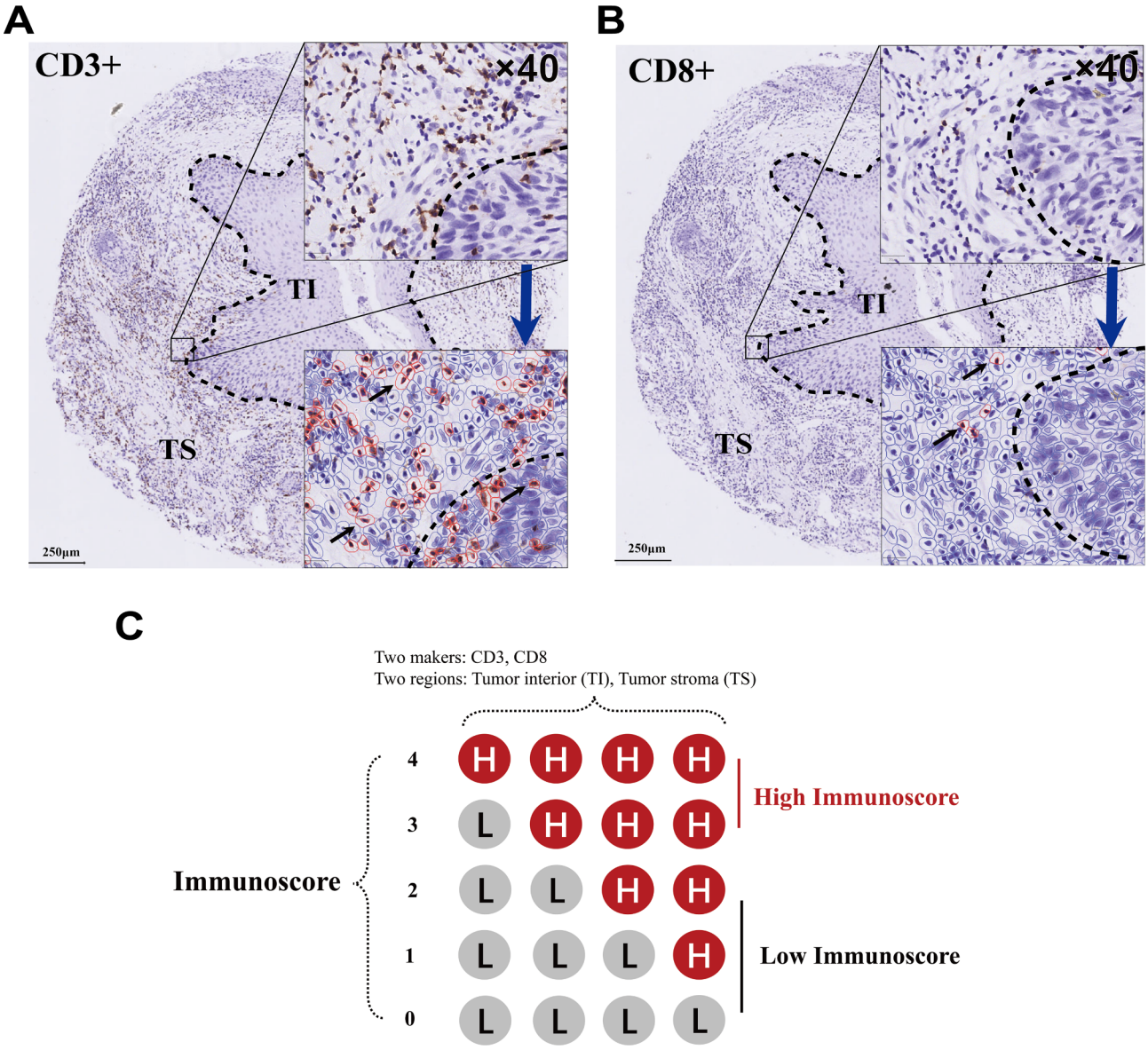

Fig. 1 The calculation scheme for immunoscore. **A** and **B** The evaluation of CD3+and CD8+T-cells present in the tumor interior (TI) and the tumor stroma (TS). **C** The immunoscore for each patient was obtained by combing four infiltration levels (two markers and two regions)

of 1.5 mm, matrix size of 512×512, and field of view of 240×240 mm^2. A 60-s delay followed intravenous injection of contrast agent (60–100 mL; ioversol and iohexol).

Tumor segmentation and habitats generation

The workflow of radiomics is shown in Fig. 2. For tumor segmentation, the open-source ITK-SNAP software (version 3.6.0) was employed. Radiologist 1 (M.Q.), with 7 years' expertise in head-and-neck MRI interpretation, manually delineated three-dimensional regions of interest (ROIs) slice by slice, ensuring whole tumor coverage. To assess reproducibility, Radiologist 1 repeated the tumor segmentation process after a minimum interval of one month. In addition, a random selection of 30 lesions was segmented by Radiologist 2 (J.R.), with 9 years' experience in head-and-neck MRI interpretation. The inter- and intra-observer reliability of radiomics features was then evaluated using intraclass correlation coefficients (ICCs).

The habitats were generated using an in-house software nnFAE, built using Python programming language. For each patient, the whole tumor ROI was divided into two subregions based on the clustering of Hounsfield unit (HU) values. The clustering was implemented at the patient level by using the Otsu threshold method. The threshold value is iteratively determined to optimize two criteria: minimizing variance within each class and maximizing variance between classes [22]. Consequently, two

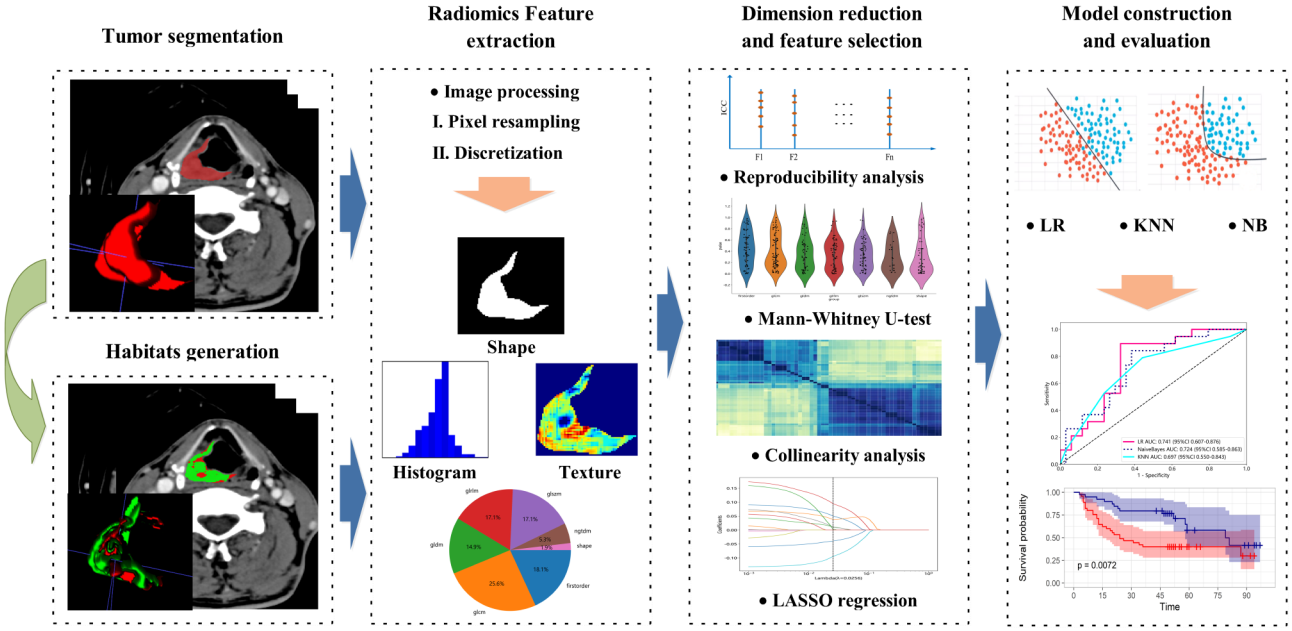

Fig. 2 Radiomics workflow and study flowchart

subregions characterized by low and high CT values (representing lacking or rich blood supply, respectively), were designated as Subregion 1 and Subregion 2, respectively.

Radiomics feature extraction

Prior to feature extraction, all images underwent two image-processing steps: (1) the standardization of in-plane resolution at 1×1 mm^2; and (2) the gray-level discretization with bin count set as 64. Subsequently, radiomics features were extracted independently from both the habitats and whole tumor using PyRadiomics (version 3.0.1). The radiomics features included four distinct categories: 14 shape- and size-based features, 17 first-order histogram features, and 75 textural features. These textural features were further categorized into five classes: gray-level co-occurrence matrix (GLCM), gray-level dependence matrix (GLDM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighboring gray-tone difference matrix (NGTDM). To capture comprehensive image information, we also utilized the images processed with Laplacian of Gaussian (LoG) filtering (values of 1, 3, and 5 mm) and wavelet transformation (four distinct combinations of high- and low-frequency bands). In total, 750 radiomics features were extracted from each ROI.

Dimension reduction and feature selection

The reproducibility of each feature was evaluated using ICC, with the following categorization: (1) ICC<0.5, poor reproducibility; (2) 0.5≤ICC<0.75, moderate reproducibility; (3) 0.75≤ICC<0.9, good reproducibility; and (4) ICC≥0.9, excellent reproducibility [23]. Features demonstrating an ICC of 0.8 or higher in both intra-observer and inter-observer analyses were deemed reliable and included in subsequent analyses. Prior to further analysis, the radiomics features underwent Z-score processing to normalize the data to a mean of 0 and variance of 1. Normalized features were then subjected to Mann–Whitney U tests. The correlation between significant features was assessed using the Spearman correlation coefficient analysis. In cases where the correlation exceeded 0.7, only the feature with the lower collinearity with the remaining features was retained. Finally, the optimal feature subsets were selected using the least absolute shrinkage and selection operator (LASSO) method, employing 10-fold cross-validation.

Model construction and evaluation

The models for predicting immunophenotyping were individually developed and validated using the training and testing cohort, respectively. Given the limited sample size, three representative and relatively straightforward classifiers were used for constructing models, including the K-nearest neighbor (KNN), logistic regression (LR), and Naive bayes (NB). Separate models were respectively built using the optimal feature subsets for the whole tumor and habitats. Based on the predictions from the best model, patients were categorized into two groups: inflamed and non-inflamed type. Furthermore, the correlation between the predicted immunophenotyping and DFS as well as OS was investigated to reinforce the clinical utility of CT radiomics.

Statistical analyses

Statistical analyses were conducted with R software (version 3.5.2; www.r-project.org). To assess the differences in clinical characteristics between patients with the inflamed and non-inflamed type, appropriate statistical tests were applied, including the chi-square test or the Mann–Whitney U test. The reproducibility of radiomics features derived from habitats was compared with those derived from whole tumor using McNemar's tests and Wilcoxon signed-rank tests. Receiver operating characteristic (ROC) curve analysis was used to evaluate the predictive capabilities of the radiomics models. The area under the curve (AUC) values of the whole tumor and habitats-based models were compared using the DeLong method. Kaplan–Meier analysis and log-rank test were used to investigate the difference in survival. Cox proportional hazards regression was employed to calculate the hazard ratio (HR) and identify independent prognostic factors. $p<0.05$ was considered to indicate statistically significant differences.

Results

Patient characteristics

The baseline characteristics of all patients are summarized in Table 1. The proportion of patients with inflamed type was 39.6% (21 out of 53) in the training cohort and 35.8% (19 out of 53) in the testing cohort. There was no significant difference in the immunophenotyping distribution between the two cohorts ($p=0.841$). Significant differences in clinical characteristics were not observed between patients with inflamed and non-inflamed type.

Comparison of reproducibility between whole tumor and habitats

Whole tumor analysis revealed that 73.1% ($n=548$) and 49.6% ($n=372$) radiomics features demonstrated good-to-excellent reproducibility in intra-observer and inter-observer assessments, respectively. Habitats analysis showed that 69.7% ($n=523$) features from Subregion 1 and 38.5% ($n=289$) features from Subregion 2 exhibited good-to-excellent reproducibility in intra-observer assessment, while 51.5% ($n=386$) features from Subregion 1 and 7.5% ($n=56$) features from Subregion 2 did so in inter-observer assessment. The ICCs of features derived from whole tumor were significantly higher than those of features derived from habitats in both intra-observer and inter-observer assessments (all $p<0.01$). Significant differences in the proportion of features exhibiting good-to-excellent reproducibility were observed only between whole tumor and Subregion 2 (all $p<0.001$). Details regarding the reproducibility of radiomics features derived from whole tumor and habitats are presented in Table 2.

Comparison of predictive performance between whole tumor and habitats

A total of 322 radiomics features from the whole tumor and 352 features (consisting of 316 from Subregion 1 and 36 from Subregion 2) from habitats exhibited satisfactory reproducibility. Significant differences were observed in 64 and 34 features from the whole tumor and Subregion 2 between patients with inflamed and non-inflamed type, respectively. However, no such differences were observed

Table 1 Clinical characteristics of patients

	Training cohort			Testing cohort		
	Non-inflamed ($n=32$)	Inflamed ($n=21$)	*p* value	Non-inflamed ($n=34$)	Inflamed ($n=19$)	*p* value
Age	61 (56, 68)	62 (62, 68)	0.176	62 (58, 65)	62 (57, 68)	0.704
Sex			1			1
Male	32 (100.00)	21 (100.00)		34 (100.00)	19 (100.00)	
Smoking			0.415			0.938
No	3 (9.38)	4 (19.05)		9 (26.47)	6 (31.58)	
Yes	29 (90.62)	17 (80.95)		25 (73.53)	13 (68.42)	
Drinking			0.144			0.313
No	12 (37.50)	13 (61.90)		17 (50.00)	13 (68.42)	
Yes	20 (62.50)	8 (38.10)		17 (50.00)	6 (31.58)	
T stage			0.987			0.845
T1-2	11 (34.38)	8 (38.10)		13 (38.24)	7 (36.84)	
T3-4	21 (65.62)	13 (61.90)		21 (61.76)	12 (63.16)	
N stage			0.738			0.667
N0	18 (56.25)	10 (47.62)		18 (52.94)	12 (63.16)	
N1-3	14 (43.75)	11 (52.38)		16 (47.06)	7 (36.84)	
TNM stage			0.887			0.803
I-II	10 (31.25)	7 (33.33)		10 (29.41)	7 (36.84)	
III-IV	22 (68.75)	14 (66.67)		24 (70.59)	12 (63.16)	

Data are expressed as median (interquartile range) or number (percentage)

Table 2 The reproducibility statuses and intra-class coefficients of whole tumor and habitats-based features

Segmentation	Analysis	Reproducibility status			ICCs	
		Poor-to-moderate	Good-to-excellent	p^1 value	Median (interquartile range)	p^2 value
Whole tumor	Intra-observer	202 (26.9%)	548 (73.1%)		0.88 (0.73–0.96)	
	Inter-observer	378 (50.4%)	372 (49.6%)		0.75 (0.53–0.89)	
Habitats-Subregion1	Intra-observer	227 (30.3%)	523 (69.7%)	0.058	0.84 (0.72–0.94)	< 0.001
	Inter-observer	364 (48.5%)	386 (51.5%)	0.304	0.76 (0.53–0.87)	0.003
Habitats-Subregion2	Intra-observer	461 (61.5%)	289 (38.5%)	< 0.001	0.66 (0.37–0.83)	< 0.001
	Inter-observer	694 (92.5%)	56 (7.5%)	< 0.001	0.28 (0.06–0.58)	< 0.001

Data are expressed as median (interquartile range) or number (percentage)

p^1 The results of the McNemar's tests comparing the proportion of radiomics features exhibiting good to excellent reproducibility between the whole tumor and the two habitats subregions

p^2 The results of Wilcoxon signed-rank tests for comparing the ICCs of radiomics features between whole tumor and two habitats subregions

ICC: intraclass correlation coefficient

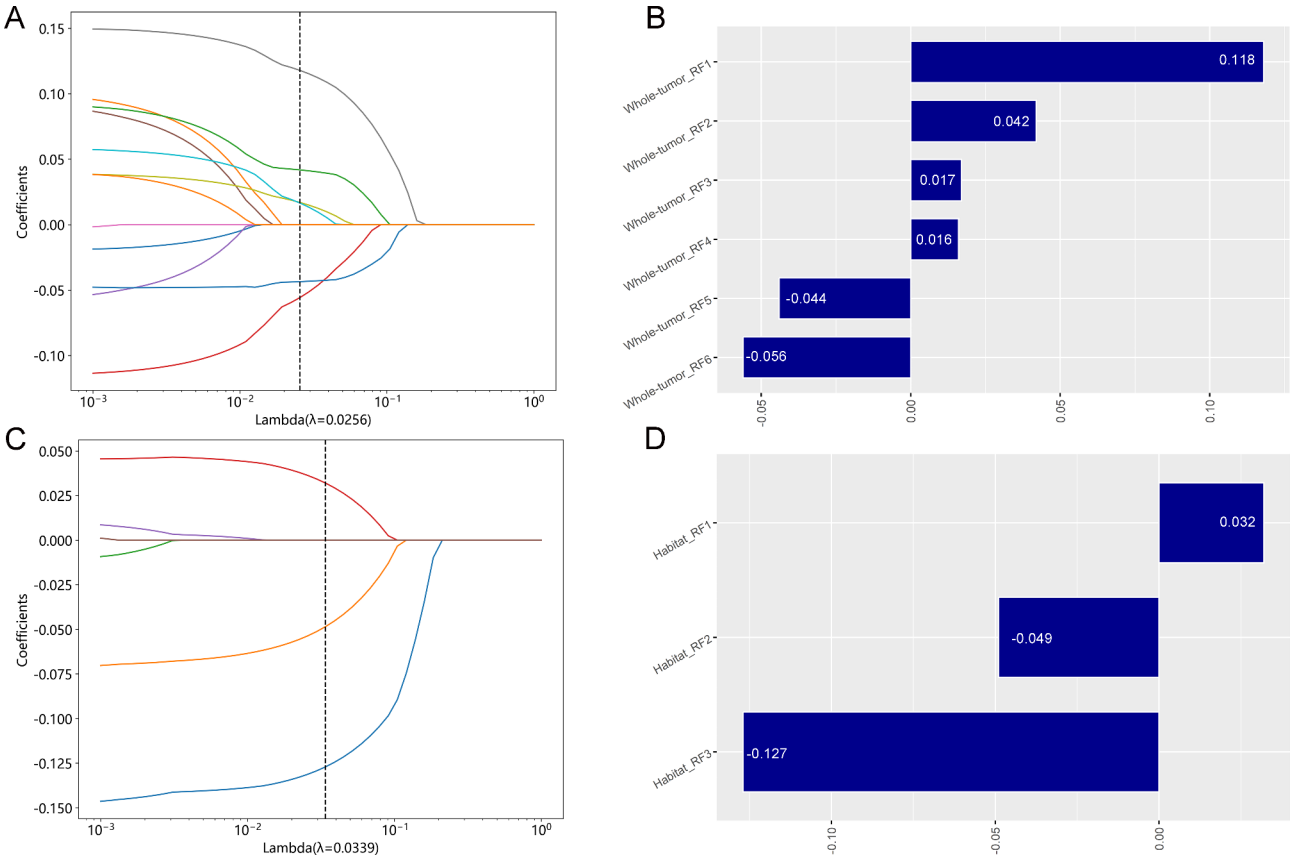

Fig. 3 The selection of optimal features for whole tumor (**A** and **B**) and habitats (**C** and **D**). **A** and **C** The least absolute shrinkage and selection (LASSO) coefficient profiles of the 12 whole tumor and 6 habitats-based features. Two dotted lines were drawn at the selected values, which respectively resulted in 6 and 3 features with non-zero coefficients. **B** and **D** The histograms show the coefficients of each optimal feature for whole tumor and habitats. Please refer to Table 3 for the code name of the selected features. RF radiomics feature

in features from Subregion 1. After conducting a collinearity analysis, 12 features from whole tumor and 6 from habitats were selected for further analysis. Finally, the LASSO regression identified 6 and 3 features with non-zero coefficients that were subsequently used to develop the whole tumor and habitats-based models, respectively (Fig. 3). The selected features for whole tumor and habitats and their ICCs are described in Table 3.

Among the whole tumor models, the LR algorithm emerged as the most predictive for immunophenotyping in the testing cohort, achieving an AUC of 0.741 and an accuracy of 0.755. The NB and KNN models achieved AUCs of 0.724 and 0.697 with accuracies of 0.698 and 0.642, respectively. However, among the habitats models, the KNN algorithm demonstrated the best performance in the testing cohort, with an AUC of 0.611 and an

Table 3 The name and intra-class coefficients of optimal feature subsets for whole tumor and habitats

Feature code	Radiomics features	ICC	
		Intra-observer	Inter-observer
Whole tumor			
RF1	Shape_Sphericity	0.95	0.85
RF2	LoG3mm_Histogram_RobustMeanAbsoluteDeviation	0.98	0.93
RF3	WaveletHH_GLCM_Correlation	0.86	0.83
RF4	WaveletHL_GLDM_SmallDependenceHighGrayLevelEmphasis	0.82	0.81
RF5	WaveletLL_GLRLM_GrayLevelNonUniformity	0.97	0.89
RF6	Log3mm_GLCM_ClusterShade	0.89	0.83
Habitats			
RF1	Subregion1_Log3mm_NGTDM_Strength	0.88	0.86
RF2	Subregion1_LoG3mm_Histogram_Energy	0.93	0.85
RF3	Subregion1_Original_Shape_MajorAxisLength	0.94	0.86

GLCM: Gray-level co-occurrence matrix, GLDM: Gray-level dependence matrix, GLRLM: Gray-level run-length matrix, ICC: Intraclass correlation coefficient, NGTDM: Neighboring gray-tone difference matrix, RF: Radiomics feature

Table 4 The performance of whole tumor and habitats-based models

ROI	Model	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Whole tumor	Training cohort						
	LR	0.781 (0.649–0.913)	0.774	0.714	0.812	0.714	0.812
	NB	0.796 (0.672–0.921)	0.774	0.810	0.750	0.680	0.857
	KNN	0.821 (0.714–0.927)	0.774	0.952	0.656	0.645	0.868
	Testing cohort						
	LR	0.741 (0.607–0.876)	0.755	0.895	0.676	0.607	0.92
	NB	0.724 (0.585–0.863)	0.698	0.842	0.618	0.552	0.875
	KNN	0.697 (0.550–0.843)	0.642	0.789	0.559	0.500	0.826
Habitats	Training cohort						
	LR	0.743 (0.607–0.878)	0.698	0.905	0.562	0.576	0.900
	NB	0.762 (0.629–0.895)	0.736	0.762	0.719	0.64	0.68
	KNN	0.857 (0.762–0.952)	0.736	0.905	0.625	0.613	0.909
	Testing cohort						
	LR	0.610 (0.451–0.769)	0.509	0.947	0.265	0.419	0.900
	NB	0.610 (0.447–0.773)	0.717	0.421	0.882	0.667	0.732
	KNN	0.611 (0.452–0.771)	0.660	0.526	0.735	0.526	0.735

Data in parentheses are 95% confidence intervals

AUC: Area under the curve, CI: Confidence interval, KNN: K nearest neighbor, LR: Logistic regression, NB: Naive Bayes, PPV: Positive predictive value, NPV: Negative predictive value

accuracy of 0.660. The LR and NB models achieved AUCs of 0.610 and 0.610, with accuracies of 0.509 and 0.717, respectively. The Delong test revealed that there was no statistically significant difference in AUCs between the best models for the whole tumor and habitats (AUC: 0.741 vs. 0.611, $p=0.112$). The performance of these radiomics models is summarized in Table 4, and the corresponding ROC curves are presented in Fig. 4. Considering the influence of classification algorithms, the comparison of performance parameters between whole tumor models and habitats models utilizing the same classifier in the testing cohort is shown in Fig. 5. The radar charts for whole tumor models exhibit larger areas compared with those of habitats models across all three classifiers.

Association between radiomics-predicted immunophenotyping and survival

The log-rank test demonstrated significantly longer survival durations among patients with inflamed type predicted by the optimal model (whole tumor model using LR) (DFS, $p=0.008$; OS, $p=0.007$) (Fig. 6). In the univariable analysis, the immunophenotyping based on the radiomics model was associated with poor DFS (HR, 0.47; $p=0.011$) and OS (HR, 0.42; $p=0.009$). Furthermore, multivariable Cox regression analysis identified the radiomics-based immunophenotyping as an independent predictor for DFS (HR, 0.46; $p=0.009$) and OS (HR, 0.42; $p=0.008$) (Table 5).

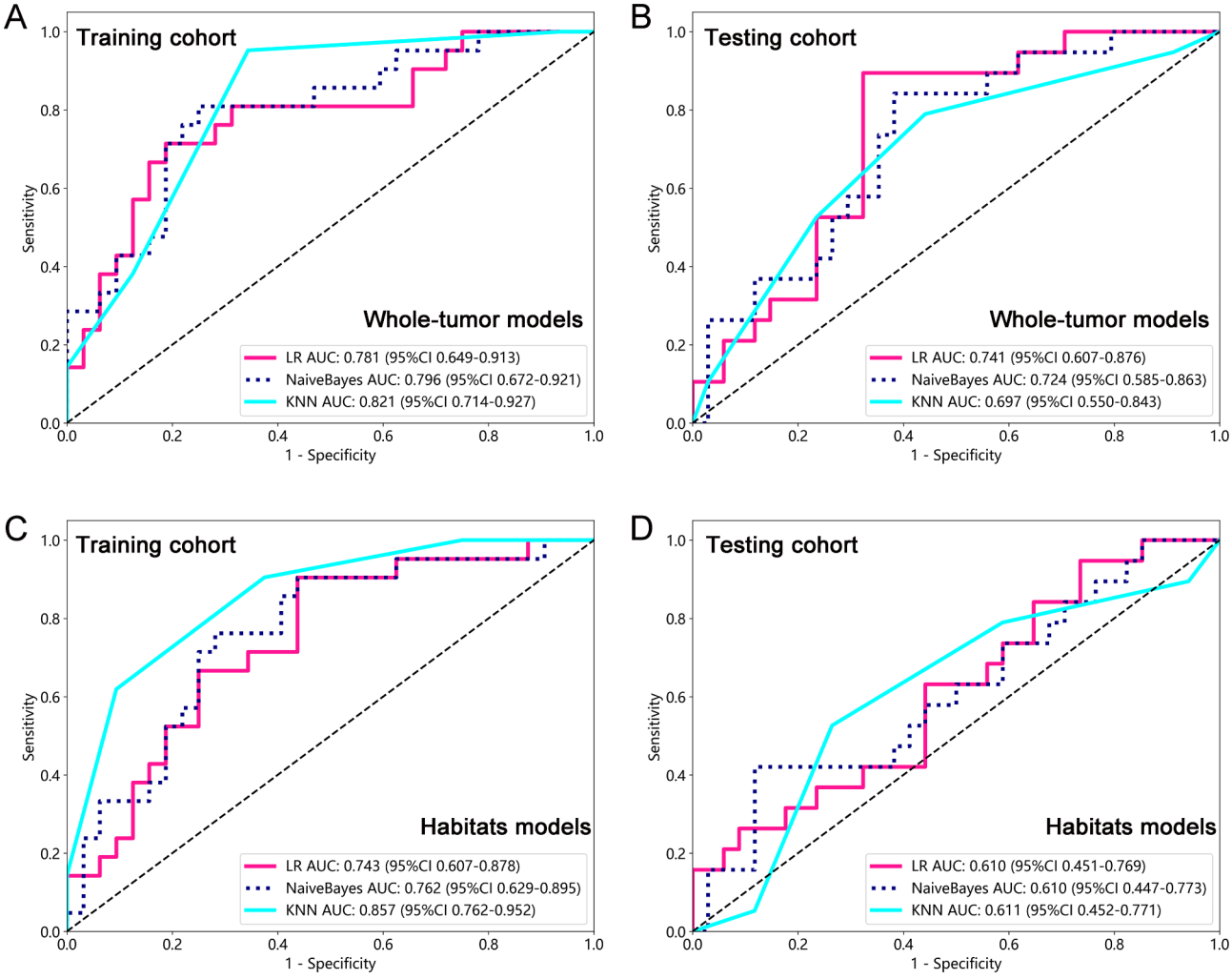

Fig. 4 Receiver operating characteristic curves of the radiomics models in the training (**A** and **C**) and testing (**B** and **D**) cohorts. **A** and **B** whole tumor models. **C** and **D** Habitats models. KNN, k-nearest neighbor; LR, logistic regression

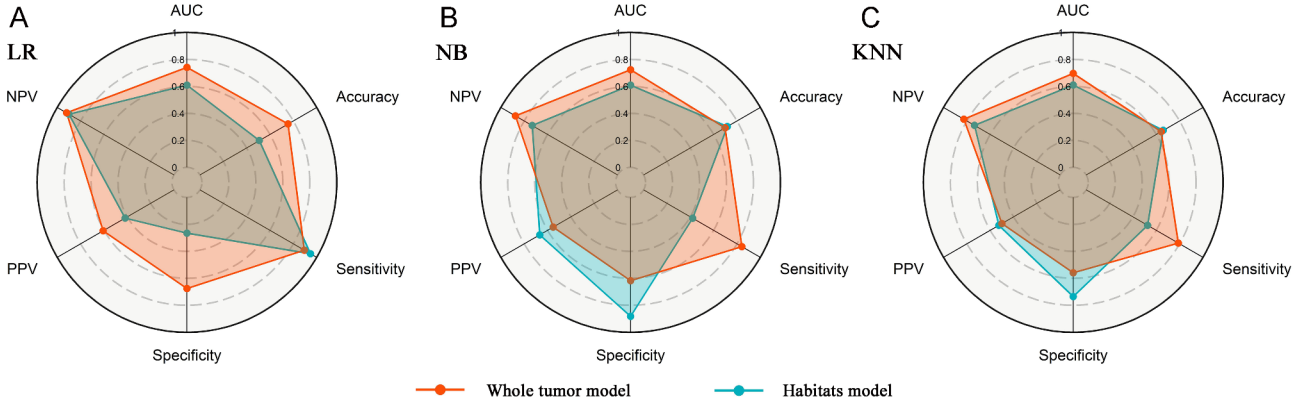

Fig. 5 Radar chart of the performance parameters of whole tumor and habitats models using different classification algorithms in the testing cohort. The radar charts for whole tumor models exhibit larger areas compared with those of habitats models across all three classifiers. **A** logistic regression (LR). **B** naive bayes (NB). **C** k-nearest neighbors (KNN)

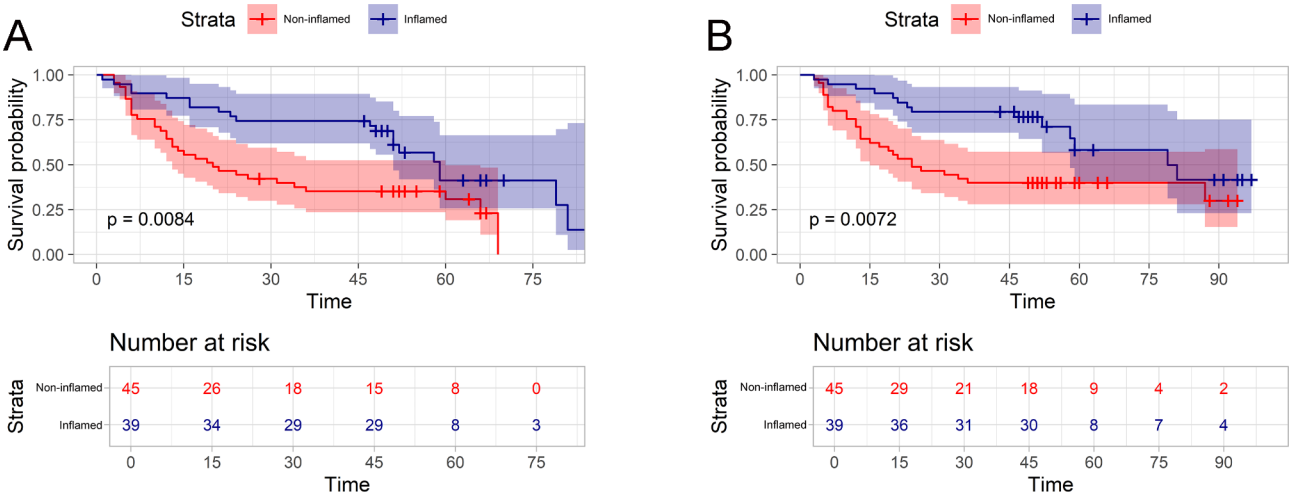

Fig. 6 Kaplan–Meier curves of immunophenotyping predicted by the optimal radiomics model. **A** Kaplan–Meier survival curves for disease-free survival. **B** Kaplan–Meier survival curves for overall survival

Table 5 Univariate and multivariate Cox analyses of risk factors of disease-free survival and overall survival

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	*p* value	HR (95% CI)	*p* value
Disease-free survival				
Age (>=62 vs. <62)	1.02 (0.58, 1.78)	0.954	—	—
Smoking (Yes vs. No)	1.49 (0.67, 3.31)	0.331	—	—
Drinking (Yes vs. No)	1.32 (0.75, 2.31)	0.332	—	—
Maximum diameter (>=32 mm vs. <32 mm)	1.92 (1.09, 3.38)	0.024	—	—
T stage (3–4 vs. 1–2)	1.55 (0.78, 3.11)	0.212	—	—
N stage (1–3 vs. 0)	2.49 (1.41, 4.41)	0.002	2.48 (1.40, 4.40)	0.002
TNM (III-IV vs. I-II)	2.49 (0.89, 6.92)	0.081	—	—
Predicted immunophenotyping (Inflamed vs. Non- inflamed)	0.47 (0.26, 0.84)	0.011	0.46 (0.26, 0.82)	0.009
Overall survival				
Age (>=62 vs. <62)	1.06 (0.58, 1.95)	0.846	—	—
Smoking (Yes vs. No)	1.36 (0.60, 3.07)	0.456	—	—
Drinking (Yes vs. No)	1.62 (0.88, 2.98)	0.125	—	—
Maximum diameter (>=32 mm vs. <32 mm)	2.19 (1.16, 4.13)	0.015	—	—
T stage (3–4 vs. 1–2)	1.21 (0.58, 2.54)	0.607	—	—
N stage (1–3 vs. 0)	2.17 (1.16, 4.03)	0.015	2.21 (1.18, 4.14)	0.013
TNM (III-IV vs. I-II)	1.92 (0.69, 5.39)	0.214	—	—
Predicted immunophenotyping (Inflamed vs. Non- inflamed)	0.42 (0.22, 0.81)	0.009	0.42 (0.22, 0.80)	0.008

Data in parentheses are 95% CIs

CI: Confidence interval, HR: Hazard ratio

Discussion

In this study, our results revealed that not only do the radiomics features derived from the whole tumor exhibit better reproducibility than those from habitats, but the whole tumor model also outperforms the habitats model for predicting immunophenotyping in the testing cohort (AUC; 0.741 vs. 0.611). Furthermore, the immunophenotyping predicted by the optimal whole tumor model was identified as an independent risk factor for unfavorable DFS (p=0.009) and OS (p=0.008) in LSCC patients. Our study showed that whole tumor-based CT radiomics is adequate for serving as a predictive biomarker of immunophenotyping and outcome prediction in LSCC patients.

In recent years, many scholars have proposed that habitats analysis can better predict tumor aggressiveness and prognosis than whole tumor radiomics, based on its enhanced ability to meticulously delineate the local heterogeneity [24–26]. Currently, both the Otsu method [22, 27] and k-means method [28, 29] have been widely used for obtaining subregions in habitats radiomics studies. The Otsu algorithm primarily utilizes the gray-level distribution of the image, rendering it less susceptible to noise and more straightforward to be interpreted.

Conversely, the k-means clustering algorithm relies on distance metrics for clustering, making it more sensitive to noise and outliers, and its results are harder to be explained. Furthermore, in comparison to k-means, Otsu's method boasts higher computational efficiency, facilitating quicker and simpler implementation. To enhance the practicality and interpretability of the habitats analysis, we utilized the straightforward Otsu method to generate two distinct subregions with varying blood supply characteristics on contrast-enhanced CT images. In radiomics research, the features from manual segmentation must be robust to generate reliable classification models. However, some habitats analysis studies have indeed overlooked the essential step of intra- and inter-observer reproducibility analysis [25, 30, 31]. In our study, we found that the ICCs of features derived from whole tumor were significantly higher than those of features derived from habitats. This is mainly because of the clustering step involved in habitats analysis, wherein even with the simple Otsu method, the resulting subregions are still susceptible to segmentation variation. Additionally, the voxel count in subregions is smaller than the whole region, further increasing the impact of segmentation variation on the stability of habitats features. Notably, the features with good-to-excellent reproducibility for Subregion 2 (high CT value) was significantly lower than that of the whole tumor, while no significant difference was observed in Subregion 1 (low CT value). We speculate that in LSCC, regions with richer blood supply tend to be distributed at the periphery of the tumor, but these regions often have unclear boundaries, leading to greater variations in their delineation than the tumor's central region with lower blood supply. Thus, despite the distinct advantages that habitats offers, it is essential to thoroughly consider potential challenges related to feature reproducibility in practical applications.

Contrary to previous findings [24–26], our study demonstrated that whole tumor-based radiomics models outperformed habitats-based models in predicting the immunophenotyping of LSCC, with maximum AUCs of 0.741 versus 0.611, respectively. However, these differences were not statistically significant ($p=0.112$), potentially due to the limited sample size. If this study had a larger sample size, the difference in performance between the two models might be clearer. Nevertheless, it cannot be denied that the habitats approach can provide more comprehensive heterogeneity information based on tumor subregions. The inferior performance of the habitats model can primarily be attributed to the exclusion of numerous radiomic features, particularly those from Subregion 2 (richer blood supply), due to poor reproducibility. We hypothesize that in future studies, when obtaining subregions derived from automated segmentation-based ROIs and avoiding the interference of interobserver variation, the habitats model may demonstrate its advantage in predicting the immunophenotyping of LSCC.

During the model-building process, we opted to use only KNN, LR, and NB classifiers, while foregoing the application of more advanced algorithms like random forests and support vector machines. KNN serves as a fundamental method for classification and regression. LR excels in classifying linear data and offers strong interpretability. NB is a simple probabilistic classifier grounded in Bayes' theorem, assuming independence among components. Considering the constraints of working with a small dataset and utilizing a limited number of features, these simpler classifiers may offer superior predictive performance and generalization capabilities. In addition, the radiomics study based on these three classifiers has been reported to be effective in predicting the aggressiveness, biological characteristics, and prognosis of head and neck cancers [32–34]. We found that the LR model based on whole tumor radiomics achieved the best performance. Similarly, radiomics models utilizing LR can be valuable tools to assess the level of TILs in head and neck cancer [33]. These results indicate that our data may be linear or linearly separable, making the LR model more effective than nonlinear approaches.

Our results further confirmed the prognostic value of CT radiomics and indicated that the predicted immunophenotyping exhibited a significant correlation with DFS and OS, and it was recognized as an independent predictor. In addition, in recent times, tumor immunotherapy has brought new hope to patients with advanced head and neck cancers. However, it is unfortunate that only 20 to 50% of tumor patients respond to immunotherapy targeting the immune checkpoint pathway [35]. Prior researches have indicated that an inflammatory tumor microenvironment characterized by high CD3+/CD8+T-cell infiltration or elevated expression of programmed death ligand-1 (PD-L1) is correlated with a patient's response to immunotherapy [36–38]. Consequently, our proposed CT radiomics model may also serve as a means to predict PD-L1 expression in LSCC and identify patients who are likely to benefit from immunotherapy.

This study has some limitations. First, given the rigorous enrollment criteria employed in this retrospective study, the dataset was relatively small for a radiomics study. Small sample size can limit the statistical power of the study, making it difficult to detect significant difference between whole-tumor models and habitats models or to generalize the findings to a larger population. Additionally, sample size limitations can introduce bias and increase the likelihood of Type II errors. Therefore, our results require further validation through multi-center and large-scale studies. Second, only two subregions based on the Otsu method were explored and hence, we

were unable to evaluate the reproducibility and predictive value of habitats features based on other methods such as k-means clustering algorithm. In future research, by employing a variety of more advanced clustering techniques, the advantages of habitats models are expected to be enhanced or more prominent. Third, only three simple classifiers were used for predictive model training; exploring more advanced machine learning algorithms could potentially further improve the results. Last, manual segmentation is not only time-consuming but also susceptible to interobserver variation. Therefore, exploring an automatic segmentation approach using deep learning techniques is paramount in the future research.

Conclusions

In summary, this study revealed that CT radiomics derived from whole tumor exhibit higher reproducibility and better performance than those from habitats for predicting immunophenotyping in LSCC. Whole tumor radiomics is sufficient for serving as a predictive biomarker of immunophenotyping and outcome prediction in LSCC patients.

Abbreviations

AUC	Area under the curve
CT	Computed tomography
DFS	Disease-free survival
ICC	Intraclass correlation coefficient
KNN	K-nearest neighbors
LSCC	Laryngeal squamous cell carcinoma
LoG	Laplacian of Gaussian
LR	Logistic regression
NB	Naive bayes
OS	Overall survival
ROC	Receiver operating characteristic
ROI	Region of interest
TIL	Tumor-infiltrating lymphocyte

Acknowledgements

We thank the OnekeyAI platform and its developers in this study.

Author contributions

M.Q. and D.Z. contributed to data collection. J.R., W.Z. and Y. S. contributed to data analysis. J.R. and Y.Y contributed to manuscript writing. J.R. and D.Z. were guarantors of integrity of the entire study and contributed to study design. All authors read and approved the final manuscript.

Funding

This work was supported by funds from the National Scientific Foundation of China (No.82101992, No.82103316 and No.82172051) and Shanghai Ninth People's Hospital (2022hbyjxys-rjl).

Data availability

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

Declarations

Ethical approval

This study was approved by the Institutional Review Board of Shanghai Eye & ENT Hospital. The informed consent was waived by the Institutional Review Board of Shanghai Eye & ENT Hospital due to retrospective study design. This study was carried out in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Radiology, Eye & ENT Hospital, Fudan University, Shanghai, China
[2]Department of Radiology, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[3]MR Scientific Marketing, Siemens Healthineers Ltd, Shanghai, China
[4]Department of Otolaryngology, Eye & ENT Hospital, Fudan University, Shanghai, China

Received: 21 August 2024 / Accepted: 5 November 2024
Published online: 11 November 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and Mortality Worldwide for 36 cancers in 185 countries. Cancer J Clin. 2021;71:209–49.
2. Siegel RL, Miller KD, Jemal A, Cancer statistics. 2018. CA: a cancer journal for clinicians. 2018;68:7–30.
3. Baird BJ, Sung CK, Beadle BM, Divi V. Treatment of early-stage laryngeal cancer: a comparison of treatment options. Oral Oncol. 2018;87:8–16.
4. Megwalu UC, Sikora AG. Survival outcomes in advanced laryngeal cancer. JAMA otolaryngology– head neck Surg. 2014;140:855–60.
5. Diakos CI, Charles KA, McMillan DC, Clarke SJ. Cancer-related inflammation and treatment effectiveness. Lancet Oncol. 2014;15:e493–503.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
7. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24:541–50.
8. Fridman WH, Zitvogel L, Sautès-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Reviews Clin Oncol. 2017;14:717–34.
9. Galon J, Lanzi A. Immunoscore and its introduction in clinical practice. Q J Nuclear Med Mol Imaging: Official Publication Italian Association Nuclear Med. 2020;64:152–61.
10. Hijazi A, Antoniotti C, Cremolini C, Galon J. Light on life: immunoscore immune-checkpoint, a predictor of immunotherapy response. Oncoimmunology. 2023;12:2243169.
11. Zhang D, Tang D, Heng Y, Zhu XK, Zhou L, Tao L, et al. Prognostic impact of Tumor-infiltrating lymphocytes in laryngeal squamous cell carcinoma patients. Laryngoscope. 2021;131:E1249–55.
12. Wang T, Zhang D, Tang D, Heng Y, Lu LM, Tao L. The role of systemic inflammatory response index (SIRI) and tumor-infiltrating lymphocytes (TILs) in the prognosis of patients with laryngeal squamous cell carcinoma. J Cancer Res Clin Oncol. 2023;149:5627–36.
13. Galon J, Bruni D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. Nat Rev Drug Discovery. 2019;18:197–218.
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278:563–77.
15. Yao Y, Jia C, Zhang H, Mou Y, Wang C, Han X, et al. Applying a nomogram based on preoperative CT to predict early recurrence of laryngeal squamous cell carcinoma after surgery. J X-Ray Sci Technol. 2023;31:435–52.
16. Zhao X, Li W, Zhang J, Tian S, Zhou Y, Xu X, et al. Radiomics analysis of CT imaging improves preoperative prediction of cervical lymph node metastasis in laryngeal squamous cell carcinoma. Eur Radiol. 2023;33:1121–31.
17. Tian R, Li Y, Jia C, Mou Y, Zhang H, Wu X, et al. Radiomics Model for Predicting TP53 Status using CT and machine Learning Approach in laryngeal squamous cell carcinoma. Front Oncol. 2022;12:823428.
18. Chen T, Li X, Mao Q, Wang Y, Li H, Wang C, et al. An artificial intelligence method to assess the tumor microenvironment with treatment outcomes for gastric cancer patients after gastrectomy. J Translational Med. 2022;20:100.

19. Han X, Cao W, Wu L, Liang C. Radiomics Assessment of the Tumor Immune Microenvironment to predict outcomes in breast Cancer. Front Immunol. 2021;12:773581.
20. Wu J, Liu W, Qiu X, Li J, Song K, Shen S, et al. A Noninvasive Approach to Evaluate Tumor Immune Microenvironment and Predict outcomes in Hepatocellular Carcinoma. Phenomics. 2023;3:549–64.
21. Wang S, Liu X, Wu Y, Jiang C, Luo Y, Tang X, et al. Habitat-based radiomics enhances the ability to predict lymphovascular space invasion in cervical cancer: a multi-center study. Front Oncol. 2023;13:1252074.
22. Liu Y, Wang P, Wang S, Zhang H, Song Y, Yan X, et al. Heterogeneity matching and IDH prediction in adult-type diffuse gliomas: a DKI-based habitat analysis. Front Oncol. 2023;13:1202170.
23. Koo TK, Li MY. A Guideline of selecting and reporting Intraclass correlation coefficients for Reliability Research. J Chiropr Med. 2016;15:155–63.
24. 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;288:26–35.
25. Hu Y, Jiang T, Wang H, Song J, Yang Z, Wang Y, et al. Ct-based subregional radiomics using hand-crafted and deep learning features for prediction of therapeutic response to anti-PD1 therapy in NSCLC. Phys Medica: PM: Int J Devoted Appl Phys Med Biology: Official J Italian Association Biomedical Phys. 2024;117:103200.
26. Wang X, Xu C, Grzegorzek M, Sun H. Habitat radiomics analysis of pet/ct imaging in high-grade serous ovarian cancer: application to Ki-67 status and progression-free survival. Front Physiol. 2022;13:948767.
27. Mu W, Liang Y, Hall LO, Tan Y, Balagurunathan Y, Wenham R, et al. (18)F-FDG PET/CT Habitat Radiomics predicts outcome of patients with cervical Cancer treated with Chemoradiotherapy. Radiol Artif Intell. 2020;2:e190218.
28. Xie C, Yang P, Zhang X, Xu L, Wang X, Li X, et al. Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. EBioMedicine. 2019;44:289–97.
29. Yuan J, Wu M, Qiu L, Xu W, Fei Y, Zhu Y, et al. Tumor habitat-based MRI features assessing early response in locally advanced nasopharyngeal carcinoma. Oral Oncol. 2024;158:106980.
30. Peng J, Zou D, Zhang X, Ma H, Han L, Yao B. A novel sub-regional radiomics model to predict immunotherapy response in non-small cell lung carcinoma. J Translational Med. 2024;22:87.
31. Zhao H, Su Y, Wang Y, Lyu Z, Xu P, Gu W, et al. Using tumor habitat-derived radiomic analysis during pretreatment (18)F-FDG PET for predicting KRAS/NRAS/BRAF mutations in colorectal cancer. Cancer Imaging: Official Publication Int Cancer Imaging Soc. 2024;24:26.
32. Tran WT, Suraweera H, Quaioit K, Cardenas D, Leong KX, Karam I, et al. Predictive quantitative ultrasound radiomic markers associated with treatment response in head and neck cancer. Future Sci OA. 2019;6:Fso433.
33. Ren J, Yang G, Song Y, Zhang C, Yuan Y. Machine learning-based MRI radiomics for assessing the level of tumor infiltrating lymphocytes in oral tongue squamous cell carcinoma: a pilot study. BMC Med Imaging. 2024;24:33.
34. Yuan Y, Ren J, Tao X. Machine learning-based MRI texture analysis to predict occult lymph node metastasis in early-stage oral tongue squamous cell carcinoma. Eur Radiol. 2021;31:6429–37.
35. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature. 2017;541:321–30.
36. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature. 2014;515:563–7.
37. Girolami I, Pantanowitz L, Barberis M, Paolino G, Brunelli M, Vigliar E, et al. Challenges facing pathologists evaluating PD-L1 in head & neck squamous cell carcinoma. J oral Pathol Medicine: Official Publication Int Association Oral Pathologists Am Acad Oral Pathol. 2021;50:864–73.
38. Paolino G, Pantanowitz L, Barresi V, Pagni F, Munari E, Moretta L, et al. PD-L1 evaluation in head and neck squamous cell carcinoma: insights regarding specimens, heterogeneity and therapy. Pathol Res Pract. 2021;226:153605.

Publisher's note