Appendix 1: 19 unique radiomics features

(I) First order features
firstorder_Entropy, firstorder_MeanAbsoluteDeviation, firstorder_Median

(II) Gray level co-occurrence matrix (GLCM) features
glcm_DifferenceAverage, glcm_DifferenceEntropy, glcm_DifferenceVariance, glcm_Imc1, glcm_Imc2, glcm_InverseVariance, glcm_JointEnergy, glcm_JointEntropy, glcm_SumEntropy

(III) Gray level run length matrix (GLRLM) features
glrlm_LongRunEmphasis, glrlm_RunEntropy, glrlm_RunVariance

(IV) Gray level size zone matrix (GLSZM) features
glszm_SizeZoneNonUniformityNormalized, glszm_SmallAreaHighGrayLevelEmphasis

(V) Neighbouring gray tone difference matrix (NGTDM) features
ngtdm_Contrast, and ngtdm_Strength

Appendix 2: habitat generation process

Our methodology for delineating tumor habitat regions was multi-faceted and involved several intricate steps:

(I) Advanced superpixel segmentation: utilizing the SLIC algorithm within the scikit-learn framework, we initially segmented each tumor's ROI into 100 subregions. The segmentation was fine-tuned with a compactness parameter set at 10.0, balancing color similarity and spatial proximity.
The SLIC algorithm is a method for superpixel segmentation. It works by clustering pixels in the image based on their color similarity and proximity in the image space. The key formula for SLIC is the distance measure, which combines color and spatial proximity:

$$D = \sqrt{d_{lab}^2 + \left(\frac{d_s}{S}\right)^2} \quad [1]$$

D is the combined distance measure.
d_{lab} is the Euclidean distance in color space (lab color space).
d_S is the Euclidean distance in the image plane.
S is the grid interval or the size of the superpixel.

(II) Comprehensive radiomic feature extraction: for each of these subregions, a detailed extraction of 107 radiomic features was performed. This included an array of shape descriptors, textural features, and first-order statistical attributes, offering a multidimensional characterization of each subregion.

(III) In-depth clustering analysis: the K-means algorithm was employed to analyze the multidimensional feature space. We experimented with varying numbers of cluster centers (ranging from 3 to 9) to identify distinct habitat regions within the tumor. The clustering performance was rigorously evaluated using the CH score, allowing us to select the most statistically significant clustering arrangement.
The K-means algorithm is a method for clustering data into K distinct clusters. The algorithm iteratively updates the centroids of each cluster to minimize the within-cluster sum of squares. The key formula for K-means is the objective function to be minimized:

$$J = \sum_{i=1}^{N} \sum_{k=1}^{K} w_{ik} \times \left\| x_i - \mu_k \right\|^2 \quad [2]$$

J is the objective function.
N is the number of data points.
K is the number of clusters.
w_{ik} is a binary indicator (1 if data point i is in cluster k, 0 otherwise).
x_i is the ith data point.
μ_k is the centroid of cluster k.
$\left\| x_i - \mu_k \right\|^2$ is the squared Euclidean distance between data point i and centroid k.

(IV) Habitat region synthesis: following the clustering analysis, subregions with identical cluster IDs were amalgamated. This synthesis resulted in the formation of comprehensive habitat regions, each representing a unique microenvironmental characteristic within the tumor.

https://dx.doi.org/10.21037/qims-24-1642

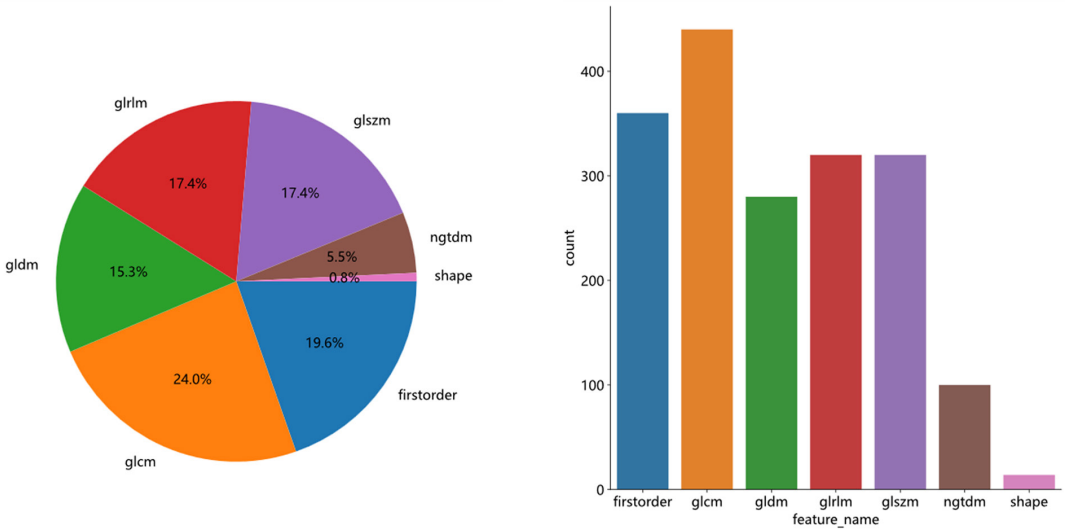

Figure S1 Number and ratio of handcrafted features.

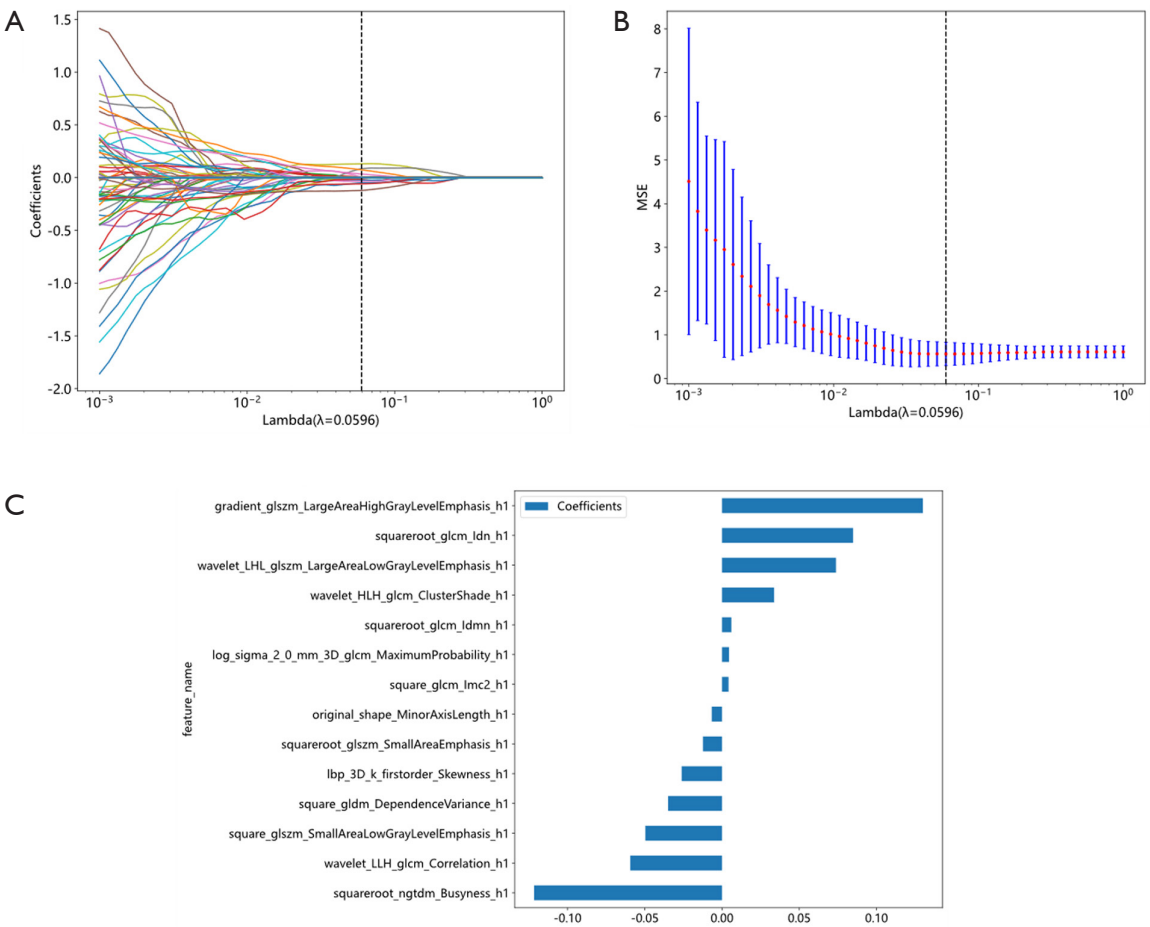

Figure S2 LASSO feature selection process. (A) Coefficients of 10-fold cross-validation. (B) MSE of 10-fold cross-validation. (C) A histogram depicting the Rad-score based on the selected features.

https://dx.doi.org/10.21037/qims-24-1642

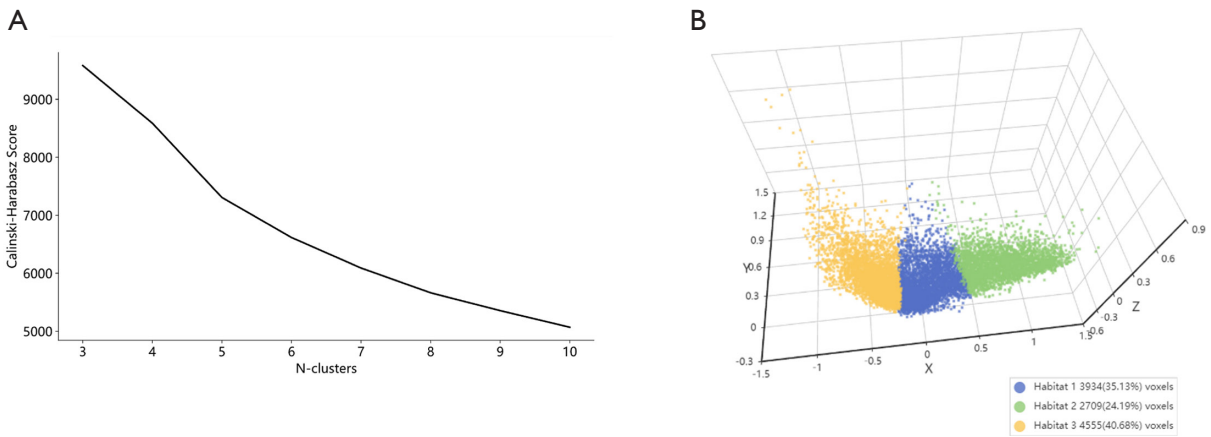

Figure S3 Cluster number selection. (A) CH scores for different clusters. (B) Visualization of cluster features. CH, Calinski-Harabasz.

Table S1 Univariable and multivariable analysis of clinical features

Feature name	OR	OR lower 95% CI	OR upper 95% CI	P value
Age	1.005	0.995	1.015	0.442
Sex	1.075	0.748	1.545	0.740
Stage	1.162	1.066	1.265	<0.05
T	0.986	0.860	1.132	0.867
N	1.203	1.055	1.372	<0.05

https://dx.doi.org/10.21037/qims-24-1642