

Machine-Learning Models Using Multidimensional Habitats MRI Features for Assessing Recurrence and Metastasis Risk in Locally Advanced Nasopharyngeal Carcinoma

Yixin Lu

Guangxi Medical University Cancer Hospital

Binling Liang

Guigang People's Hospital

Xiaoyi Zeng

Guangxi Medical University Cancer Hospital

Shilai Zhang

Guangxi Medical University Cancer Hospital

Zhenyu Chen

Guangxi Medical University Cancer Hospital

Jie Chen

Guangxi Medical University Cancer Hospital

Hai Liao

42442427@qq.com

Guangxi Medical University Cancer Hospital

Research Article

Keywords: Locally advanced nasopharyngeal carcinoma, Radiomics, intratumoral heterogeneity, Multidimensional Habitats, machine learning models

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10588615/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Approximately 30%–40% of patients with locally advanced nasopharyngeal carcinoma (LANPC) develop local recurrence or distant metastasis. Early identification of high-risk patients is therefore important for improving prognosis.

Methods

A total of 1,285 patients with LANPC from three medical centers were enrolled. Patients from center A were randomly divided into a training set ($n = 693$) and a validation set ($n = 297$) at a ratio of 7:3, while patients from centers B and C were combined as an external test set ($n = 295$). Ten independent machine learning models and 26 stacked ensemble models were developed. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, precision, and F1 score. The incremental value of multidimensional habitat (MH) features over clinical features alone was assessed using integrated discrimination improvement (IDI), net reclassification improvement (NRI), calibration curves, and decision curve analysis (DCA). The SHapley Additive exPlanations (SHAP) method was used to interpret the final model.

Results

After recursive feature elimination (RFE), four selected clinical features, namely N category, treatment, platelet count (PLT), and smoking status, were identified and further combined with two imaging features, namely multidimensional habitats derived from contrast-enhanced T1-weighted imaging (T1C MH) and multidimensional habitats derived from T2-weighted imaging (T2WI MH), for model development. The stacked ensemble model combining support vector machine (SVM), gradient boosting machine (GBM), and logistic regression (LR) achieved the best overall performance, with a mean area under the receiver operating characteristic curve (AUC) of 0.761 (95% confidence interval [CI]: 0.719–0.798) in the 5-fold cross-validation set, 0.742 (95% CI: 0.679–0.801) in the validation set, and 0.721 (95% CI: 0.660–0.776) in the external test set.

Conclusion

In conclusion, the stacking model constructed based on SVM, GBM, and LR, integrating MH features and clinical features, demonstrated favorable predictive performance for recurrence and metastasis in LANPC patients, and holds promise as an effective tool for individualized risk assessment.

Evidence Level: 3

Technical Efficacy: Stage 2

Introduction

Nasopharyngeal carcinoma (NPC) is a malignant tumor originating from the mucosal epithelium of the nasopharynx, with a predilection for the pharyngeal recess (fossa of Rosenmüller) (1). According to the 2022 global cancer statistics, 120,416 new cases of NPC and 73,476 deaths occurred worldwide in 2022 (2). Approximately 70% of new cases of NPC occur in East and Southeast Asia (1). Approximately 70% of cases of NPC are diagnosed at a locally advanced stage (3).

Although current medical approaches such as induction chemotherapy, concurrent chemoradiotherapy (CCRT), adjuvant chemotherapy, and targeted therapy can prolong progression-free survival in most patients with NPC (4, 5), 30–40% of patients experience local recurrence or distant metastasis (6). Recurrence and metastasis are the main causes of treatment failure in NPC. Patients with recurrent or metastatic NPC (RM-NPC) have a poor prognosis, with a median overall survival of less than 20 months (7). In the era of precision medicine, developing effective methods for risk assessment and personalized treatment strategies for high-risk patients is important for improving the prognosis of patients with NPC in less-developed regions, particularly in East and Southeast Asia.

MRI, with its excellent soft tissue contrast and radiation-free properties, has been integrated into the management workflow of NPC, including lesion detection and diagnosis, TNM staging, radiotherapy planning, and evaluation of the response to treatment (1, 8). The advent of big-data mining and artificial intelligence has revealed that conventional MRI contain vast amounts of underutilized information, particularly the unique textural patterns and intratumoral heterogeneity (ITH) features within lesions (9). Radiomics is an emerging approach that can decode and transform digitally encrypted medical information reflecting tumor pathophysiological characteristics into mineable high-dimensional data. These radiomic features not only reveal tumor phenotypic traits but also provide insights into the tumor microenvironment. Such data can complement clinical indicators, treatment strategies, and genomic profiles to establish a precise and reliable clinical decision-support system (10). Previous studies have demonstrated the potential of radiomic features for predicting response to induction chemotherapy and prognostic evaluation in NPC (11, 12). However, most radiomics research has been conducted under the assumption that the entire tumor mass constitutes a single homogeneous tissue or system (11–13). Research has confirmed the existence of multiple genetic subclones within tumors, which can be regarded as distinct ecological communities or habitats of tumor cells, each exhibiting unique genetic profiles and biological behaviors (12). Li et al(14) proposed an unsupervised method to assess tumor heterogeneity using the spatial-distribution patterns of radiomic features and clustering analysis. This approach has shown promising results in predicting the prognosis of non-small cell lung cancer and evaluation of neoadjuvant chemotherapy response in breast cancer (14–17). Building on previous research on ITH in breast cancer and tumor habitats in liver and NPC (18–20), we developed a framework for evaluating tumor habitats. Our novel approach used a habitat classification scheme based on three key feature categories: (1) shape features (sphericity, surface-to-volume ratio), (2) density features (mean entropy, kurtosis), and (3) textural parameters were selected as clustering cores, including gray-level co-occurrence matrix (GLCM) contrast, GLCM joint entropy, gray-level size zone matrix (GLSZM) zone size non-uniformity, gray-level run length matrix (GLRLM) run length non-uniformity, and GLRLM short run emphasis (21). The methodology enhances the biological characterization of

tumor heterogeneity through cluster analysis based on shape, density, and texture features, thereby improving the model's capacity to assess recurrence and metastasis risks in patients with locally advanced NPC (LANPC).

We developed a machine-learning Multidimensional Habitats analysis method for comprehensive quantification of radiomic features in LANPC. Using patient cohorts from three institutions, we systematically compared the abilities of conventional radiomics, the ITH score, and Multidimensional Habitats features to assess 5-year recurrence and metastasis following treatment in patients with LANPC, and simultaneously validated the cross-platform generalizability of conventional radiomic features, ITH score, and Multidimensional Habitats characteristics.

Materials and Methods

Study Population

This study was conducted in accordance with the Declaration of Helsinki and approved by the institutional review boards of all participating centers. The requirement for informed consent was waived owing to the retrospective design and use of anonymized data. The results were reported according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines and the study was classified as a Type III hybrid effectiveness-implementation study (22). The set included 1496 patients treated between January 2014 and April 2020 at three medical centers (A, B, and C), of whom 486, and 232 patients from Centers A and C, respectively, received 3.0-T MRI; and 658 and 120 patients from Centers A and B, respectively, received 1.5-T MRI.

The final patient selection was based on predefined inclusion and exclusion criteria. The inclusion criteria were: (i) Pathologically confirmed NPC; (ii) Locally advanced disease (Stage II–III) according to the 9th edition of the Union for International Cancer Control/American Joint Committee on Cancer (UICC/AJCC) staging system (8); (iii) Availability of a pretreatment MRI performed within 1 month prior to initiating therapy; and (iv) Complete pretreatment clinical records and hematologic test results.

The exclusion criteria were: (i) Treatment discontinuation due to patient refusal or intolerable side effects; (ii) History of a prior malignancy; (iii) Receipt of antitumor therapy before MRI examination; (iv) Pretreatment MRI images that could not be assessed or failure of feature extraction; and (v) Patients with severe comorbidities (Fig. 1).

The set from Center A was randomly split into training and validation sets in a 7:3 ratio, and the sets from Centers B and C were combined to form an external test set for model evaluation.

MRI Examination and Preprocessing Procedures

The specific MRI scanning protocols and parameters (including manufacturer, magnetic field strength, contrast agent usage, and detailed sequence information) for each participating center are provided in **Supplementary Material 1**.

Treatment Protocols and Follow-up

The primary treatment regimens were categorized into four distinct approaches: (1) CCRT, (2) induction chemotherapy followed by CCRT (IC+CCRT), (3) CCRT followed by adjuvant chemotherapy (CCRT + AC), and (4) induction chemotherapy with CCRT and adjuvant chemotherapy (IC+CCRT + AC) **Supplementary Material 2**.

Patients underwent standardized follow-up assessments according to the following schedule: every 1–3 months during the initial two years post-treatment, biannually for the subsequent two years, and annually thereafter. The follow-up duration for these patients was 5 years, with tumor recurrence and metastasis confirmed by radiological examination or pathological diagnosis.

Tumor Segmentation

The overall workflow is shown in Fig. 2. Two board-certified head and neck radiologists (A and B, with 7 and 8 years of specialized experience, respectively) independently performed three-dimensional tumor delineation on T1C sequences using ITK-SNAP software (version 3.8.0; <http://www.itksnap.org/pmwiki/pmwiki.php>). In cases of uncertain 3D segmentation boundaries, a final determination was made by a senior radiologist with 17 years of experience. To ensure the reliability of segmentation results, a random sampling method was employed to assess inter-observer consistency, with quantitative evaluation performed using Dice similarity coefficients (DSC) for region of interest (ROI) delineation.

Image Preprocessing

Standardized preprocessing was performed on MRI sets from all three centers, encompassing both 1.5-T and 3.0-T field strengths. The pipeline included: (1) image denoising using discrete wavelet transformation(23), (2) N4 bias field correction to mitigate magnetic field inhomogeneity(24), (3) isotropic resampling to 1mm×1mm×1mm voxel dimensions via linear interpolation(25), and (4) affine registration of T1C, T2WI, and corresponding masks with manual refinement for misaligned cases(26).

Feature Extraction and Heterogeneity Analysis

Conventional Radiomics Feature Extraction

Radiomic feature extraction was performed using PyRadiomics software (version 3.1; <https://pyradiomics.readthedocs.io>). Features were extracted from both the original images and derived images processed with LoG filtering, wavelet transformation, and square/square root/logarithm/exponential/gradient transformations. A total of seven categories of features were extracted: (1) morphological features (shape), (2) first-order statistical features, (3) gray-level co-occurrence matrix (GLCM), (4) gray-level run-length matrix (GLRLM), (5) gray-level size zone matrix (GLSZM), (6) gray-level dependence matrix (GLDM), and (7) neighboring gray-tone difference matrix (NGTDM). In total, 1670 radiomic features were extracted from each region of interest.

Calculation of the ITH score

The tumor is partitioned into distinct "tissue habitats" through extraction and clustering of pixel-level radiomic features using unsupervised learning methods. The spatial fragmentation and dispersion of these habitats are then quantified by the ITH score formula:

$$1 - \frac{1}{S_{total}} \sum_{i=1}^v \frac{S_{i,max}}{n_i}$$

ultimately yielding a comprehensive score that measures ITH.

Multidimensional Habitats Analysis Method.

The tumor habitat analysis was performed using the following computational pipeline: First, the existing 3D tumor masks were partitioned into 100 super-voxels through the Simple Linear Iterative Clustering (SLIC) algorithm(18). Three distinct feature categories served as clustering cores: (1) Shape features (Sphericity, surface-to-volume ratio), (2) density features (Mean, Entropy, Kurtosis), and (3) texture features (GLCM contrast, GLCM joint entropy, GLSZM zone size non-uniformity, GLRLM run length non-uniformity, and GLRLM short run emphasis). Gaussian Mixture Model clustering was then applied, with the optimal cluster number (ranging from 2–10) determined through Bayesian Information Criterion (BIC) optimization. This approach enabled biologically meaningful re-segmentation of the 3D masks into distinct tumor habitats.

Radiomic Feature Selection and Model Construction

Batch Effect Correction and Feature Screening

To reduce batch effects across centers, we standardized sets from different centers using Z-score normalization. Feature selection followed a two-step procedure: (1) initial screening using F-tests to identify the top 500 statistically significant features, followed by (2) refined selection via the minimum Redundancy Maximum Relevance (mRMR) algorithm to derive the optimal 100 features.

LASSO Regression-Based Feature Optimization

The pre-selected feature subset was subsequently processed using Least Absolute Shrinkage and Selection Operator (LASSO) regression within a 5-fold cross-validation framework. The model automatically optimized the regularization strength parameter (alpha) to compress coefficients of non-informative features to zero, thereby achieving automated feature selection. This approach not only reduced model complexity and improved interpretability but also effectively minimized overfitting risks. The final model construction incorporated only features with non-zero coefficients identified through LASSO, generating both conventional radiomics and Multidimensional Habitats-based radiomics scores (Supplementary Table 1).

Machine-Learning Model Development and Evaluation

Clinical features Selection

The study included the following clinical features in the analysis: age, sex, smoking history, Epstein-Barr virus (EBV) DNA status, pathological type, T category, N category, treatment regimen, height, weight; and the following laboratory parameters: hemoglobin level (Hb), platelet count (PLT), lymphocyte count (LYM), and albumin level (ALB).

Independent Machine Learning Model Development

We constructed a VIF-screened clinical feature set comprising 14 clinical features. Recursive feature elimination (RFE) was used to rank the features, and model performance was evaluated stepwise from the full feature set down to a single feature. The models were implemented in Python 3.9 and included 10 algorithms: logistic regression (LR), k-nearest neighbors (KNN), decision tree (DT), random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), TabPFN, XGBoost, LightGBM, and CatBoost. All models were optimized on the training set using grid search with stratified 5-fold cross-validation. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, precision, F1 score, sensitivity, and specificity.

Stacking Ensemble Model Development

From models showing no evidence of overfitting in both training and validation sets, the top five were selected as base learners for stacking, by assessing their comprehensive ranking in the validation set. Multiple stacking combinations were explored, with hyperparameter tuning conducted via stratified 5-fold cross-validation and grid search to enhance robustness, reduce overfitting, and optimize parameters. Stacking ensembles were evaluated using the same metrics as the base models.

Clinical Benefit of MH Features

Using the same evaluation framework, we compared “MH feature + Clinical feature” against “Clinical feature,” assessing discriminative performance via ROC curves. Clinical utility was quantified by decision curve analysis (DCA), and the incremental value of MH features over clinical models was assessed using the net reclassification improvement (NRI) and integrated discrimination improvement (IDI).

SHAP Interpretability Framework

We applied SHAP (Shapley Additive Explanations) to the optimal model to explain and visualize how predictors contribute to recurrence and metastasis risk, addressing the black-box nature of ML models. SHAP offers both global and local interpretability: globally ranking feature importance and direction of effect, and locally detailing patient-level risk drivers and contribution magnitudes. This improves model transparency and supports clinical decision-making and personalized risk assessment.

Statistical Analysis

Data were summarized using descriptive statistics: normally distributed continuous variables as mean $\pm$ standard deviation, and non-normally distributed continuous variables as median (interquartile range, IQR). Missing data were handled using Multiple Imputation by Chained Equations (MICE). Clinical variables were screened with the variance inflation factor (VIF) to control multicollinearity, and retained accordingly.

For assessing recurrence and metastasis risk, model performance was primarily evaluated by a composite ranking score on the independent validation set. Specifically, each model received dense, descending ranks across AUC, accuracy, precision, F1 score, sensitivity, and specificity; the average of these ranks constituted the composite score, with lower values indicating superior overall performance.

Results

After excluding 154 cases from the internal set and 57 cases from the external test set according to the exclusion criteria, this study included a total of 990 cases in the internal set and 295 cases in the external test set. The internal set was divided into a training set ($n = 693$) and a validation set ($n = 297$) (Fig. 1). The baseline demographic and clinical features of each group are summarized in Table 1. Analysis of 100 random MRI samples by two head and neck radiologists showed DSC all greater than 0.8 (**Supplementary Fig. 1**).

Table 1
Baseline Table.

Characteristic	Training set	Validation set	External test set	P Value1	P Value2
Sample Size	693	297	295		
Age(y)	45.000(36.000–52.000)	45.000(37.000–53.000)	49.000(40.000–56.000)	0.308	< 0.001
Height(mm)	165.000(160.000–170.000)	165.000(159.000–170.000)	164.000(157.000–168.000)	0.718	0.001
Weight(kg)	60.000(53.000–67.000)	59.000(53.000–66.000)	58.000(52.000–66.000)	0.264	0.092
Hb(g/L)	138.000(125.000–149.000)	139.000(127.000–150.000)	136.435(124.000–148.000)	0.185	0.273
PLT (10 ⁹ /L)	270.000(231.000–323.000)	272.000(229.950–329.000)	267.000(223.500–313.000)	0.916	0.177
LYM(10 ⁹ /L)	1.810(1.440–2.240)	1.790(1.470–2.260)	1.740(1.415–2.180)	0.909	0.103
ALB (g/L)	40.200(37.900–42.600)	40.100(37.900–42.600)	42.300(40.000–45.400)	0.868	< 0.001
Sex				1	0.582
Female	180 (25.974%)	77 (25.926%)	82 (27.797%)		
Male	513 (74.026%)	220 (74.074%)	213 (72.203%)		
Smoking				0.088	< 0.001
No	409 (59.019%)	193 (64.983%)	267 (90.508%)		
Yes	284 (40.981%)	104 (35.017%)	28 (9.492%)		
EBV DNA (copies/ml)				0.668	< 0.001
<500	428 (61.760%)	188 (63.300%)	135 (45.763%)		
≥500	265 (38.240%)	109 (36.700%)	160 (54.237%)		
Pathological				0.121	0.036
WHO 1–2	69 (9.957%)	40 (13.468%)	17 (5.763%)		
WHO 3	624 (90.043%)	257 (86.532%)	278 (94.237%)		
Note: P value 1 indicates the statistical comparison between the training set and the Validation set. P value 2 indicates the statistical comparison between the training set and the external test set.					

Characteristic	Training set	Validation set	External test set	P Value1	P Value2
T Category(%)				0.049	< 0.001
T1	17 (2.453%)	2 (0.673%)	11 (3.729%)		
T2	213 (30.736%)	74 (24.916%)	55 (18.644%)		
T3	259 (37.374%)	128 (43.098%)	108 (36.610%)		
T4	204 (29.437%)	93 (31.313%)	121 (41.017%)		
N Category(%)				0.05	< 0.001
N0	17 (2.453%)	3 (1.010%)	2 (0.678%)		
N1	148 (21.356%)	81 (27.273%)	32 (10.847%)		
N2	315 (45.455%)	139 (46.801%)	186 (63.051%)		
N3	213 (30.736%)	74 (24.916%)	75 (25.424%)		
Treatment				0.31	< 0.001
CCRT	200 (28.860%)	94 (31.650%)	48 (16.271%)		
IC+CCRT	451 (65.079%)	193 (64.983%)	202 (68.475%)		
CCRT + AC	26 (3.752%)	7 (2.357%)	11 (3.729%)		
IC+CCRT + AC	16 (2.309%)	3 (1.010%)	34 (11.525%)		
<i>Note: P value 1 indicates the statistical comparison between the training set and the Validation set. P value 2 indicates the statistical comparison between the training set and the external test set.</i>					

Feature selection

After VIF (**Supplementary Table 2**) screening and recursive feature elimination (RFE), four clinical features were ultimately retained: N category, treatment, platelet count (PLT), and smoking status (Fig. 3).

Performance evaluation of independent ML models: Based on the 6-feature set, 10 standalone ML models were evaluated. On the validation set, SVM achieved the best overall performance, ranking first in the comprehensive evaluation, with the highest AUC of 0.745 (95% confidence interval [CI]: 0.679–0.802) and the highest accuracy of 0.721. TabPFN ranked second overall and showed the highest F1 score (0.609) and sensitivity (0.755), with an AUC of 0.742 (95% CI: 0.676–0.804). In the 5-fold cross-validation analysis, GBM achieved the best comprehensive ranking, whereas TabPFN and SVM yielded the highest AUCs of 0.763 (95% CI: 0.722–0.801) and 0.760 (95% CI: 0.720–0.794), respectively. On the training set,

most models showed improved discrimination, with DT, RF, and LightGBM achieving AUCs above 0.9, whereas KNN showed an apparently perfect AUC but extremely unbalanced sensitivity and specificity. The comprehensive ranking results of the models on the validation set are presented in Tables 2 and 3. Detailed performance metrics for the training set and the 5-fold cross-validation are provided in Supplementary Tables 3,4 and 5, respectively. Additionally, the ROC curves and calibration curves for both the training and validation sets are illustrated in Supplementary Figs. 2 and 3, respectively, with further 5-fold cross-validation results depicted in Supplementary Fig. 4.

Table 2
Performance of the 10 machine learning models on the validation set.

Model	AUC (95%CI)	ACC	PRE	F1	SEN	SPE
SVM	0.745 (0.679–0.802)	0.721	0.55	0.591	0.638	0.759
TabPFN	0.742 (0.676–0.804)	0.694	0.511	0.609	0.755	0.665
CatBoost	0.728 (0.662–0.790)	0.7	0.525	0.544	0.564	0.764
GBM	0.723 (0.660–0.788)	0.697	0.518	0.563	0.617	0.734
LR	0.726 (0.655–0.781)	0.677	0.493	0.583	0.713	0.66
RF	0.711 (0.645–0.776)	0.687	0.506	0.475	0.447	0.798
XGBoost	0.697 (0.633–0.766)	0.687	0.506	0.475	0.447	0.798
LightGBM	0.697 (0.632–0.765)	0.673	0.481	0.446	0.415	0.793
KNN	0.670 (0.604–0.744)	0.316	0.316	0.481	1	0
DT	0.613 (0.543–0.680)	0.646	0.452	0.498	0.553	0.69

Table 3

Comprehensive ranking of the 10 base machine learning models on the internal validation set.

Model	AUC Rank	ACC Rank	PRE Rank	F1 Rank	SEN Rank	SPE Rank	Composite Score
SVM	1	1	1	2	4	4	2.167
TabPFN	2	4	4	1	2	7	3.333
CatBoost	3	2	2	5	6	3	3.5
GBM	5	3	3	4	5	5	4.167
LR	4	6	6	3	3	8	5
RF	6	5	5	8	8	1	5.5
XGBoost	7	5	5	8	8	1	5.667
LightGBM	8	7	7	9	9	2	7
KNN	9	9	9	7	1	9	7.333
DT	10	8	8	6	7	6	7.5

Stacked model development and evaluation

Based on the composite scores of the base learners, five leading algorithms, namely SVM, GBM, LR, TabPFN, and CatBoost, were used to construct 26 stacking configurations. On the validation set, SVM + GBM + LR achieved the best composite ranking, with an AUC of 0.742 (95% CI, 0.679–0.801), accuracy of 0.734, Precision of 0.565, F1 score of 0.622, sensitivity of 0.691, and specificity of 0.754. Although SVM + TabPFN + GBM + LR yielded the highest validation AUC of 0.747 (95% CI, 0.680–0.806), its overall composite rank was only 11. In the 5-fold cross-validation analysis, TabPFN + GBM + LR achieved the best composite ranking, whereas SVM + GBM + LR ranked seven with a mean AUC of 0.761 (95% CI, 0.719–0.798) (Fig. 4, Supplementary Tables 6–9). Considering both discrimination and ranking stability across sets, SVM + GBM + LR was selected as the final stacked model. DeLong tests showed that the selected ensemble significantly outperformed LR, GBM, RF, XGBoost, LightGBM, KNN, and DT, but not SVM, TabPFN, or CatBoost (Table 4; Supplementary Tables 10 and 11).

Table 4
Comparison of AUC performance between base models and stacking ensemble models.

Stacking Model	Base Learner	AUC of Stacking Model (95%CI)	AUC of Base Learner (95% CI)	Δ AUC	P value
SVM + GBM+ LR	SVM	0.742 (0.679–0.801)	0.745 (0.679–0.802)	-0.003	0.717
SVM + GBM+ LR	TabPFN	0.742 (0.679–0.801)	0.742 (0.676–0.804)	0.000	0.498
SVM + GBM+ LR	CatBoost	0.742 (0.679–0.801)	0.728 (0.662–0.790)	0.015	0.058
SVM + GBM+ LR	GBM	0.742 (0.679–0.801)	0.723 (0.660–0.788)	0.019	0.044
SVM + GBM+ LR	LR	0.742 (0.679–0.801)	0.726 (0.655–0.781)	0.016	0.028
SVM + GBM+ LR	RF	0.742 (0.679–0.801)	0.711 (0.645–0.776)	0.031	0.012
SVM + GBM+ LR	XGBoost	0.742 (0.679–0.801)	0.697 (0.633–0.766)	0.045	0.003
SVM + GBM+ LR	LightGBM	0.742 (0.679–0.801)	0.697 (0.632–0.765)	0.046	0.009
SVM + GBM+ LR	KNN	0.742 (0.679–0.801)	0.670 (0.604–0.744)	0.072	0.002
SVM + GBM+ LR	DT	0.742 (0.679–0.801)	0.613 (0.543–0.680)	0.129	0.000

Incremental value of MH scores:

NRI and IDI analyses showed that adding MH features improved the predictive performance of the baseline Clin model. On the validation set, T1C MH + T2WI MH + Clin achieved the largest overall improvement over the Clin model (NRI, 0.793; IDI, 0.022), followed by T1C MH + Clin (NRI, 0.766; IDI, 0.019), whereas T2WI MH + Clin showed a smaller gain (NRI, 0.432; IDI, 0.007) (**Supplementary Fig. 5**). In the ROC analysis, T1C MH + Clin achieved the highest AUC on the validation set (0.743; 95% CI, 0.680–0.802), followed closely by T1C MH + T2WI MH + Clin (0.742; 95% CI, 0.679–0.801), whereas T2WI MH + Clin and Clin alone showed lower AUCs of 0.679 (95% CI, 0.613–0.740) and 0.556 (95% CI, 0.480–0.627), respectively. Overall, T1C MH contributed most to performance improvement, whereas T2WI MH provided only limited incremental benefit (Supplementary Figs. 6–8).

External validation

Based on the horizontal comparison of feature sets, the integrated MH + Clin feature set was selected for external testing. In addition to Clinical alone, ITH + Clin and Rad + Clin were included as benchmark

feature sets. In the external test set, MH + Clin showed the best performance, achieving an AUC of 0.721(95% CI,0.660–0.776), compared with 0.644(95% CI,0.578–0.705) for Rad + Clin, 0.594(95% CI,0.524–0.662) for ITH + Clin, and 0.463(95% CI,0.338–0.538) for Clinical alone (Fig. 5, Supplementary Table 12).

Model Interpretation and Visualization

At the meta-learner level, SHAP analysis indicated that the stacked model integrated contributions from SVM, GBM, and LR. SHAP analysis identified T1C MH as the most important feature in the final model, followed by T2WI MH, N category, treatment, platelet count (PLT), and smoking status. The overall SHAP summary and feature-importance plots showed that T1C MH made the largest contribution to risk prediction in patients with locally advanced nasopharyngeal carcinoma (LANPC). Representative waterfall plots further demonstrated the contribution of each feature to the predicted probability at the individual-patient level (Fig. 6).

Discussion

This study innovatively employed ten base models and twenty-six ensemble models to systematically evaluate the risk of recurrence and metastasis in patients with LANPC. The SVM + GBM + LR ensemble showed good performance in both the 5-fold cross-validation set and the independent validation set. In addition, MH features effectively improved the predictive power of the clinical feature set.

Predicting patient outcomes is critical in clinical practice, and feature selection is central to machine learning model development because it directly affects predictive accuracy and clinical applicability(27). However, there is currently a lack of guidelines on retaining optimal feature sets and feature types in predictive models(28). Existing multi-machine-learning studies predominantly rely on clinical experience and conventional feature selection methods to determine the retained feature sets. This study included ten base learners covering mainstream modeling paradigms, including linear models, tree-based models, support vector machines, randomized ensembles, gradient boosting methods, and a deep learning-based model for tabular data. As shown by the RFE results, when the original 14 clinical features were progressively reduced, the average performance of the 10 machine-learning models in 5-fold cross-validation and the independent validation set became more stable, and the fluctuation amplitude, as reflected by the standard deviation, generally decreased, suggesting improved robustness after removal of redundant or noisy variables. After constructing the simplified Clin + MH model, generalization performance improved substantially across models. Among the base learners, TabPFN achieved the highest AUC in 5-fold cross-validation. On the independent validation set, SVM achieved the highest AUC and the best overall ranking, with TabPFN ranking second and LR, GBM, and CatBoost also showing competitive performance. Notably, SVM, LR, and TabPFN exhibited relatively small train-to-validation AUC gaps, suggesting stronger robustness after feature simplification, whereas DT and KNN remained unstable, and RF, LightGBM, and XGBoost still showed comparatively larger declines from the training to the validation set. These findings suggest that parsimonious feature selection not only improved

discrimination but also enhanced model stability and reduced performance variability among the stronger base learners. Importantly, each model has its own inductive biases and error structures, adapting differently to data distributions, feature scales, and degrees of nonlinearity. Therefore, a single model has inherent limitations: for example, the Gini/MDI-based importance measure in random forests exhibits systematic bias toward continuous and high-cardinality categorical features, with the bias becoming more pronounced under strong feature correlations; logistic regression, on the other hand, is susceptible to multicollinearity(29–31).

To enhance the stability of single learners and reduce their biases, this study adopted a stacking framework, hierarchically integrating multiple base classifiers with a meta-learner. The raw data were first processed by different base learners to generate predictions or intermediate representations, after which the meta-classifier synthesized the inputs and outputs of the base models to learn their weights, achieving cross-paradigm integration on the same set(32). Predictions and features from each layer were combined across continuous hierarchical levels, forming new and more robust predictions. Previous studies have shown that stacked models can reveal latent patterns and complex associations between risk factors and symptoms, fully leveraging the strengths of each base model, mitigating the impact of data heterogeneity, and more effectively capturing nonlinear relationships(33).

Compared with conventional radiomics and a previously established spatial-distribution–based ITH score (14), MH features—which quantify intratumoral heterogeneity—demonstrate stronger associations with recurrence and metastasis in LANPC and superior generalizability in independent external test set. When integrated with conventional clinical variables, MH substantially improves model reclassification and discrimination (as evidenced by increases in NRI and IDI), indicating that it provides independent and complementary information beyond clinical risk factors. Notably, T1C MH is the principal contributor to these gains, while adding T2WI MH yields only modest additional improvement. Collectively, MH features constitute a promising imaging biomarker with clear translational potential, enabling more precise risk stratification, optimization of surveillance strategies, and individualized therapeutic decision-making.

we performed Multidimensional habitat decomposition of the primary lesion on contrast-enhanced MRI, delineating three classes of subregions: Density habitats ($S_0 \rightarrow S_3$), which characterize the spatial transition from low to high density; Shape habitats ($S_0 \rightarrow S_1$), which capture the morphological evolution from regular to irregular tumor contours; and Texture habitats ($S_0 \rightarrow S_2$), which describe the progression from homogeneous tissue to highly heterogeneous texture complexity. Centered on these subregions, we constructed a composite MH index. Across both base learners and the stacking model, MH was consistently identified as the most critical risk factor associated with LANPC recurrence and metastasis. The index ultimately incorporated 35 subregional features (9 shape, 17 texture, 9 density) derived from both T1C and T2WI modalities. These span six feature families: first-order features, reflecting overall signal intensity and its distribution; GLCM, describing spatial dependence and directionality between pixels; GLRLM, characterizing gray-level run structures; GLSZM, quantifying non-uniformity of zone size and gray levels; GLDM, representing gray-level dependence; and NGTDM, capturing spatial frequency

changes and local heterogeneity. Specifically, high entropy is widely regarded as an imaging marker of intratumoral heterogeneity and is associated with greater aggressiveness, biological instability, and unfavorable prognosis(34, 35); Concurrently, the prominent representation of low percentiles and minimum values highlights distinct but complementary aggressive subregions within the two modalities. In T1C, low values typically reflect low perfusion, cystic changes, or hypoxic necrotic components(36). Conversely, in T2WI, lower signal intensities often correspond to areas of high cellular density, elevated nucleocytoplasmic ratio, or dense fibrotic stroma(37). In GLCM, higher correlation denotes more ordered texture; ClusterShade and the maximal correlation coefficient (MCC) capture asymmetry and complex dependencies, and elevated values often indicate microstructural complexification and increased heterogeneity(38). In GLRLM, higher ShortRunLowGrayLevelEmphasis reflects a greater prevalence of short runs with low gray levels, suggesting patches of low perfusion/necrosis and micro-heterogeneity (38, 39). In GLSZM, increased GLN/GLNN and Size-Zone Non-Uniformity indicate a more uneven distribution of gray levels across zones of different sizes, reflecting a “patchy” morphological pattern (26, 38). In GLDM, higher LowGrayLevelEmphasis reflects a larger proportion of low-gray-level dependent pixels, while LargeDependenceHighGrayLevelEmphasis delineates regions with contiguous high-signal intensities. (38, 40). Furthermore, the inclusion of NGTDM features captures abrupt spatial changes in voxel intensities, further reinforcing the characterization of aggressive tumor phenotypes(41).

A higher N stage signals wider nodal involvement and greater aggressiveness, with a correspondingly increased risk of distant metastasis(42, 43). In addition, studies have reported that both elevated and reduced PLT are associated with worse prognosis and a higher risk of metastasis in NPC(44, 45). Furthermore, smoking history has been identified as a critical risk factor; Ouyang et al. demonstrated that compared to never-smokers, both former and current smokers face significantly elevated risks of mortality, disease progression, locoregional recurrence, and distant metastasis(46). In this model, the treatment regimen contributes weakly and is included primarily as a covariate for adjustment to improve robustness and calibration; its long-term impact across different clinical stages warrants further evaluation within stratified populations and regimen-specific analyses.

Notably, this study emphasized the interpretability and potential clinical utility of the risk-assessment model for assessing recurrence and metastasis in LANPC. Although current conventional radiomics combined with machine-learning models have demonstrated potential in evaluating tumor treatment response and prognosis (11, 12, 47), most of these studies remain “black-box” in nature and lack validation using external test sets. To address this limitation, we used the SHAP method to visually demonstrate the importance and contribution of each feature to the model prediction. The force plots further enabled dynamic visualization of feature importance distributions for individual patients with different risk levels, transforming an otherwise “black-box” NPC model into a transparent and interpretable tool.

Study Limitations

This study has several limitations: (1) Nasopharyngeal carcinoma tumor ROI delineation relied primarily on manual segmentation, highlighting the need to develop fully automated segmentation techniques; (2) Retrospectively acquired features may be influenced by inter-center variations in instrumentation, scanning parameters, contrast administration, and magnetic field strength, potentially affecting result reproducibility and stability - although we have minimized multicenter bias through batch effect correction; (3) Future studies should expand sample sizes, control variables, and incorporate multi-omics data (genomics, pathomics, etc.) to further validate the clinical value of combined clinical and Multidimensional Habitats features for LANPC recurrence/metastasis risk assessment.

Conclusions

The MH+Clin models integrating clinical and Multidimensional Habitats features demonstrate accurate cross-center assessment performance for recurrence/metastasis risk in LANPC patients. This modeling approach establishes a new paradigm for nasopharyngeal carcinoma feature extraction, providing a robust tool for personalized treatment and risk stratification in locally advanced disease.

Declarations

Compliance with Ethical Standards

Guarantor:

The scientific guarantor of this publication is Guangxi Medical University Cancer Hospital.

Conflict of Interest:

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and Biometry:

Python 3.9 kindly provided statistical advice for this manuscript.

One of the authors has significant statistical expertise.

No complex statistical methods were necessary for this paper.

Informed Consent:

Written informed consent was not required for this study because retrospective investigation.

Ethical Approval:

The need for informed consent was waived by the Institutional Review Boards

because of the retrospective investigation.

Study subjects or cohorts overlap:

This study was not registered and study protocols were not published before this study..

Methodology

Methodology: retrospective

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form (including individual details, images, or videos). All authors have reviewed and approved the final version of the manuscript and consent to its publication in Cancer Imaging.

Ethics approval and consent to participate

This retrospective study was approved by the Institutional Review Boards of the following institutions: Guangxi Medical University Cancer Hospital (approval No. KY2026826), Wuzhou Red Cross Hospital (approval No. LL2025-112-01), and Guigang People's Hospital (approval No. E2026-018-01). The need for written informed consent was waived due to the retrospective nature of the investigation. The original ethics approval documents are available for verification by the editorial office upon reasonable request, but are not publicly disclosed to protect institutional data security.

Funding

This work was funded by the Guangxi Science and Technology Program (Grant No. AB23026018), the National Natural Science Foundation of China (Grant No. 82460338), the Guangxi Natural Science Foundation (Grant No. 2023GXNSFAA026249), and the Young Scholars Fund of Guangxi Medical University (Grant No. GXMUYSFB2026078).

Author Contribution

Y.L., B.L., and X.Z. contributed equally to this work and share first authorship. Y.L., B.L., and X.Z. performed the data collection and analysis. S.Z. and Z.C. contributed to the interpretation of imaging data. J.C. and H.L. conceived and designed the study, and were major contributors in writing and revising the manuscript. J.C. and H.L. share corresponding authorship. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to institutional data protection policies and ethical restrictions, as they were collected from internal multicenter data sources. However, they are available from the corresponding author upon reasonable request, subject to institutional review and approval.

References

1. Chen Y-P, Chan ATC, Le Q-T, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet*. 2019;394(10192):64–80. 10.1016/s0140-6736(19)30956-0.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J Clin*. 2024;74(3):229–63. 10.3322/caac.21834.
3. Lai P, Chen X, Pei W, Huang X, Fang Z, Yang F, Ying Y, Song Y, Jin W, Lu S, Lu B, Liao H. Development and validation of an MRI radiomics-based interpretable machine learning model for predicting the progression-free survival in locally advanced nasopharyngeal carcinoma. *Quant Imaging Med Surg*. 2025;15(6):5347–61. 10.21037/qims-24-1860.
4. Wang W, Peng S, Wu H, Luo Y, Yuan F, Lin Z, Cheng G, Chen S. Association of tumor downstaging after neoadjuvant chemotherapy with survival in patients with locally advanced nasopharyngeal carcinoma: a retrospective cohort study. *J Cancer Res Clin Oncol*. 2021;147(10):2913–22. 10.1007/s00432-021-03690-8.
5. Setakornnukul J, Thephamongkhol K. Neoadjuvant chemotherapy followed by concurrent chemoradiotherapy versus concurrent chemoradiotherapy followed by adjuvant chemotherapy in locally advanced nasopharyngeal carcinoma. *BMC Cancer*. 2018;18(1). 10.1186/s12885-018-4210-3.
6. Anthony TC, Chan EPH, Roger KC, Ngan, Stewart Y, Tung ACK, Cheng WT, Ng, Victor HF, Lee, Brigitte BY, Ma HC, Cheng, Frank CS, Wong, Herbert HF, Loong M, Tong, Darren MC, Poon AT, Ahuja AD, King K, Wang F, Mo, Benny CY, Zee KC, Allen Chan YM, Dennis Lo. Analysis of Plasma Epstein-Barr Virus DNA in Nasopharyngeal Cancer After Chemoradiation to Identify High-Risk Patients for Adjuvant Chemotherapy: A Randomized Controlled Trial. *J Clin Oncol*. 2018;JCO2018777847. 10.1200/JCO.2018.77.7847.
7. Wang FHWX, Feng J, Li Q, Xu N, Hu XC, Liao W, Jiang Y, Lin XY, Zhang QY, Yuan XL, Huang HX, Chen Y, Dai GH, Shi JH, Shen L, Yang SJ, Shu YQ, Liu YP, Wang W, Wu H, Feng H, Yao S, Xu RH. Efficacy, Safety, and Correlative Biomarkers of Toripalimab in Previously Treated Recurrent or Metastatic

- Nasopharyngeal Carcinoma: A Phase II Clinical Trial (POLARIS-02). *J Clin Oncol*. 2021;39(7):704–12. 10.1200/JCO.20.
8. Pan J-J, Mai H-Q, Ng WT, Hu C-S, Li J-G, Chen X-Z, Chow JCH, Wong E, Lee V, Ma L-Y, Guo Q-J, Liu Q, Liu L-Z, Xu T-T, Gong X-C, Qiang M-Y, Au K-H, Liu T-C, Chiang CL, Xiao Y-P, Lin S-J, Chen Y-B, Guo S-S, Wong CHL, Tang L-Q, Xu Z-Y, Jia Y-Z, Peng W-S, Hu L-P, Lu T-Z, Jiang F, Cao C-N, Xu W, Ma J, Blanchard P, Williams M, Glastonbury CM, King AD, Patel SG, Seethala RR, Colevas AD, Fan D-M, Chua MLK, Huang SH, O’Sullivan B, Lydiatt W, Lee AWM. Ninth Version of the AJCC and UICC Nasopharyngeal Cancer TNM Staging Classification. *JAMA Oncol*. 2024;10(12):1627. 10.1001/jamaoncol.2024.4354.
 9. Ng WT, Yuen KT, Au KH, Chan OSH, Lee AWM. Staging of nasopharyngeal carcinoma – The past, the present and the future. *Oral Oncol*. 2014;50(6):549–54. 10.1016/j.oraloncology.2013.06.003.
 10. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Reviews Clin Oncol*. 2017;14(12):749–62. 10.1038/nrclinonc.2017.141.
 11. Deng M, Lin Y, Yan L, Chen C, Fei Z, Ding J. A bibliometric analysis of nasopharyngeal carcinoma radiomics: trends and insights. *Front Oncol*. 2025;15. 10.3389/fonc.2025.1506778.
 12. Suryani L, Lee HPY, Teo WK, Chin ZK, Loh KS, Tay JK. Precision Medicine for Nasopharyngeal Cancer –A. *Rev Curr Prognostic Strategies Cancers*. 2024;16(5):918. 10.3390/cancers16050918.
 13. Gatenby RAGO, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology*. 2013;269(1):8–15. 10.1148/radiol.13122697.
 14. Li J, Qiu Z, Zhang C, Chen S, Wang M, Meng Q, Lu H, Wei L, Lv H, Zhong W, Zhang X. ITHscore: comprehensive quantification of intra-tumor heterogeneity in NSCLC by multi-scale radiomic features. *Eur Radiol*. 2022;33(2):893–903. 10.1007/s00330-022-09055-0.
 15. Huang Y, Wang X, Cao Y, Lan X, Hu X, Mou F, Chen H, Gong X, Li L, Tang S, Wang L, Zhang J. Nomogram for Predicting Neoadjuvant Chemotherapy Response in Breast Cancer Using MRI-based Intratumoral Heterogeneity Quantification. *Radiology*. 2025;315(1). 10.1148/radiol.241805.
 16. Qi H, Zuo Z, Lin S, Chen Y, Li H, Hu D, Zhou Y, Qi W, Chen H. Assessment of intratumor heterogeneity for preoperatively predicting the invasiveness of pulmonary adenocarcinomas manifesting as pure ground-glass nodules. *Quant Imaging Med Surg*. 2025;15(1):272–86. 10.21037/qims-24-734.
 17. Xu J, Miao J-g, Wang C-x, Zhu Y-p, Liu K, Qin S-y, Chen H-s, Lang N. CT-based quantification of intratumoral heterogeneity for predicting distant metastasis in retroperitoneal sarcoma. *Insights into Imaging*. 2025;16(1). 10.1186/s13244-025-01977-9.
 18. Shi Z, Huang X, Cheng Z, Xu Z, Lin H, Liu C, Chen X, Liu C, Liang C, Lu C, Cui Y, Han C, Qu J, Shen J, Liu Z. MRI-based Quantification of Intratumoral Heterogeneity for Predicting Treatment Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology*. 2023;308(1). 10.1148/radiol.222830.

19. Yuan J, Wu M, Qiu L, Xu W, Fei Y, Zhu Y, Shi K, Li Y, Luo J, Ding Z, Sun X, Zhou S. Tumor habitat-based MRI features assessing early response in locally advanced nasopharyngeal carcinoma. *Oral Oncol.* 2024;158:106980. 10.1016/j.oraloncology.2024.106980.
20. Zhuo L, Chen W, Xing L, Li X, Song Z, Dong J, Zhang Y, Li H, Cui J, Han Y, Hao J, Wang J, Yin X, Li C. MRI-based quantification of intratumoral heterogeneity for intrahepatic mass-forming cholangiocarcinoma grading: a multicenter study. *Insights into Imaging.* 2025;16(1). 10.1186/s13244-025-01985-9.
21. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin J-C, Pieper S, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104–7. 10.1158/0008-5472.can-17-0339.
22. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med.* 2015;162(1):W1–73. 10.7326/m14-0698.
23. Mishro PK, Agrawal S, Panda R, Abraham A. A Survey on State-of-the-Art Denoising Techniques for Brain Magnetic Resonance Images. *IEEE Rev Biomed Eng.* 2022;15:184–99. 10.1109/rbme.2021.3055556.
24. Tustison NJ, Avants BB, Cook PA, Yuanjie Z, Egan A, Yushkevich PA, Gee JC. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging.* 2010;29(6):1310–20. 10.1109/tmi.2010.2046908.
25. Nyúl LGUJ, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging.* 2000;19(2):143–50. 10.1109/42.836373.
26. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—how-to guide and critical reflection. *Insights into Imaging.* 2020;11(1). 10.1186/s13244-020-00887-2.
27. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. *BMJ.* 2024;386:e078276. 10.1136/bmj-2023-078276.
28. Hu J, Xu J, Li M, Jiang Z, Mao J, Feng L, Miao K, Li H, Chen J, Bai Z, Li X, Lu G, Li Y. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. *eClinicalMedicine.* 2024;68:102409. 10.1016/j.eclinm.2023.102409.
29. Hollmann N, Müller S, Purucker L, Krishnakumar A, Körfer M, Hoo SB, Schirrmeister RT, Hutter F. Accurate predictions on small data with a tabular foundation model. *Nature.* 2025;637(8045):319–26. 10.1038/s41586-024-08328-6.
30. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics.* 2007;8(1). 10.1186/1471-2105-8-25.
31. Heidema AG, Boer JMA, Nagelkerke N, Mariman ECM, van der Feskens ADL. The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. *BMC Genet.* 2006;7(1). 10.1186/1471-2156-7-23.

32. Zhang J, Zhu H, Chen Y, Yang C, Cheng H, Li Y, Zhong W, Wang F. Ensemble machine learning approach for screening of coronary heart disease based on echocardiography and risk factors. *BMC Med Inf Decis Mak.* 2021;21(1). 10.1186/s12911-021-01535-5.
33. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble Learning for Disease Prediction. *Rev Healthc.* 2023;11(12):1808. 10.3390/healthcare11121808.
34. Zhang H, Graham CM, Elci O, Griswold ME, Zhang X, Khan MA, Pitman K, Caudell JJ, Hamilton RD, Ganeshan B, Smith AD. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology.* 2013;269(3):801–9. 10.1148/radiol.13130110/-/DC1.
35. Gillies RJ, Kinahan PE, Radiomics HH. Images Are More than Pictures. They Are Data *Radiol.* 2016;278(2):563–77.
36. Yang F, Wei H, Li X, Yu X, Zhao Y, Li L, Li Y, Xie L, Wang S, Lin M. Pretreatment synthetic magnetic resonance imaging predicts disease progression in nonmetastatic nasopharyngeal carcinoma after intensity modulation radiation therapy. *Insights into Imaging.* 2023;14(1). 10.1186/s13244-023-01411-y.
37. Martucci M, Russo R, Schimperia F, D'Apolito G, Panfili M, Grimaldi A, Perna A, Ferranti AM, Varcasia G, Giordano C, Gaudino S. Magnetic Resonance Imaging of Primary Adult Brain Tumors: State of the Art and Future Perspectives. *Biomedicines.* 2023;11(2):364. 10.3390/biomedicines11020364.
38. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook GJR, Davatzikos C, Depeursinge A, Desseroit M-C, Dinapoli N, Dinh CV, Echegaray S, El Naqa I, Fedorov AY, Gatta R, Gillies RJ, Goh V, Götz M, Guckenberger M, Ha SM, Hatt M, Isensee F, Lambin P, Leger S, Leijenaar RTH, Lenkowicz J, Lippert F, Losnegård A, Maier-Hein KH, Morin O, Müller H, Napel S, Nioche C, Orhac F, Pati S, Pfaehler EAG, Rahmim A, Rao AUK, Scherer J, Siddique MM, Sijtsma NM, Socarras Fernandez J, Spezi E, Steenbakkers RJHM, Tanadini-Lang S, Thorwarth D, Troost EGC, Upadhaya T, Valentini V, van Dijk LV, van Griethuysen J, van Velden FHP, Whybra P, Richter C, Löck S. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology.* 2020;295(2):328–38. 10.1148/radiol.2020191145.
39. Galloway MM. Texture analysis using gray level run lengths. *Comput Graphics Image Process.* 1975;4(2):172–9. [https://doi.org/10.1016/S0146-664X\(75\)80008-6](https://doi.org/10.1016/S0146-664X(75)80008-6).
40. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing.* 1983;23(3):341–352. [https://doi.org/10.1016/0734-189X\(83\)90032-4](https://doi.org/10.1016/0734-189X(83)90032-4)
41. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybernetics.* 1989;19(5):1264–74. 10.1109/21.44046.
42. Huang C-L, Guo R, Li J-Y, Xu C, Mao Y-P, Tian L, Lin A-H, Sun Y, Ma J. Tang L-I. Nasopharyngeal carcinoma treated with intensity-modulated radiotherapy: clinical outcomes and patterns of failure

- among subsets of 8th AJCC stage IVa. Eur Radiol. 2019;30(2):816–22. 10.1007/s00330-019-06500-5.
43. Guo R, Mao YP, Tang LL, Chen L, Sun Y, Ma J. The evolution of nasopharyngeal carcinoma staging. Br J Radiol. 2019;92(1102):20190244.
 44. Chen Y-P, Chen C, Mai Z-Y, Gao JIN, Shen L-J, Zhao B-C, Chen M-K, Chen G, Yan F, Huang T-Y, Xia Y-F. Pretreatment platelet count as a predictor for survival and distant metastasis in nasopharyngeal carcinoma patients. Oncol Lett. 2015;9(3):1458–66. 10.3892/ol.2015.2872.
 45. Yang D, Li P, Meng Z, Hu X, Huang Z, Huang H, Dong H, Qin Y, Chen C, Chen X, Li Z, Zhou Z, Li Y, Kang M. Combined pretreatment neutrophil-lymphocyte ratio and platelet-lymphocyte ratio predicts survival and prognosis in patients with non-metastatic nasopharyngeal carcinoma: a retrospective study. Sci Rep. 2024;14(1). 10.1038/s41598-024-59131-2.
 46. OuYang P-Y, Su Z, Mao Y-P, Liang X-X, Liu Q, Deng W, Xie F-Y. Prognostic Impact of Cigarette Smoking on the Survival of Patients with Established Nasopharyngeal Carcinoma. Cancer Epidemiol Biomarkers Prev. 2013;22(12):2285–94. 10.1158/1055-9965.epi-13-0546.
 47. Zhao L, Gong J, Xi Y, Xu M, Li C, Kang X, Yin Y, Qin W, Yin H, Shi M. MRI-based radiomics nomogram may predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma. Eur Radiol. 2019;30(1):537–46. 10.1007/s00330-019-06211-x.

Figures

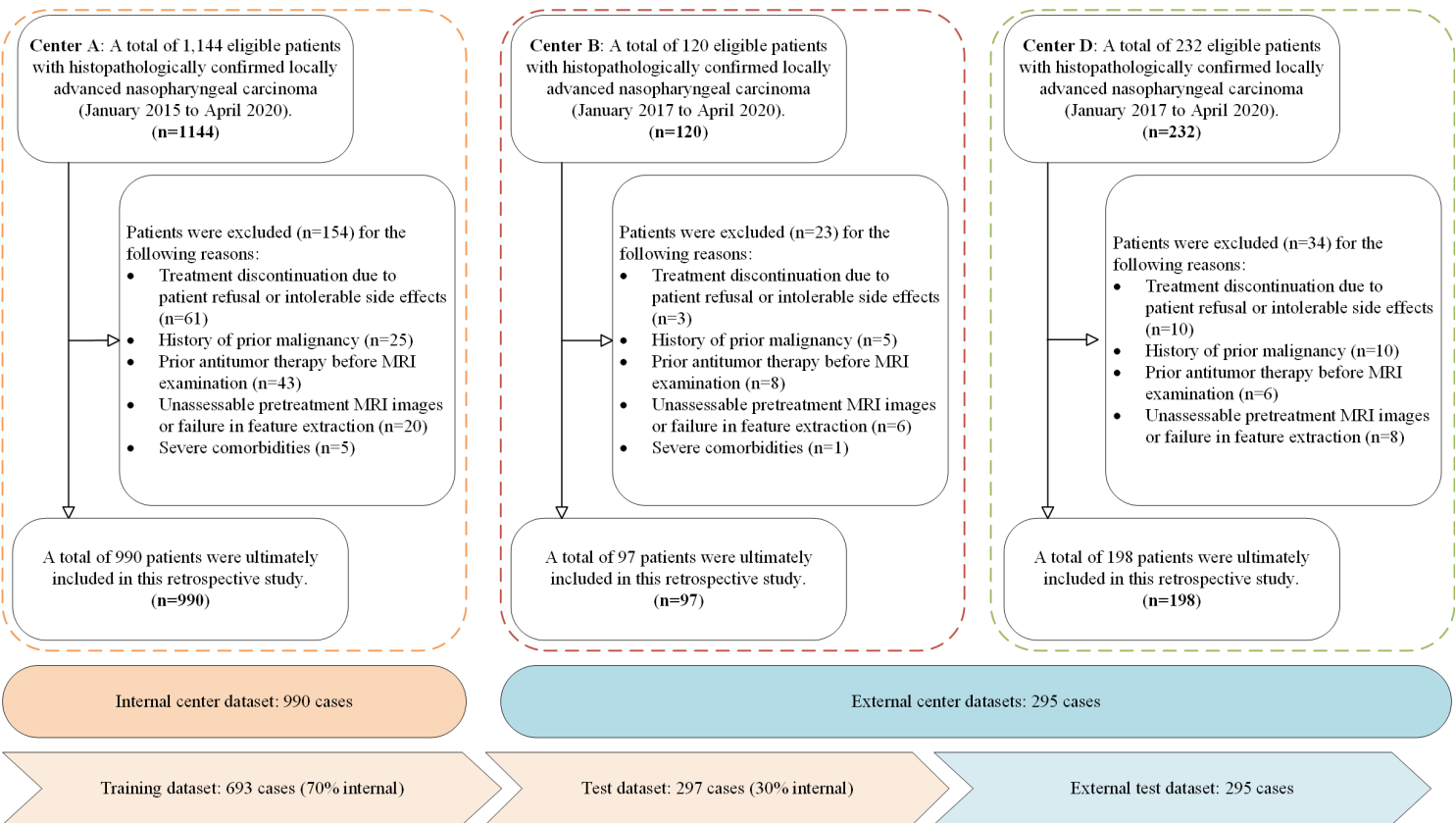


Figure 1

Study Design and Patient Enrollment. This study included a total of 1285 patients with LANPC from Three participating hospitals.

Abbreviations: LANPC, locally advanced nasopharyngeal carcinoma.

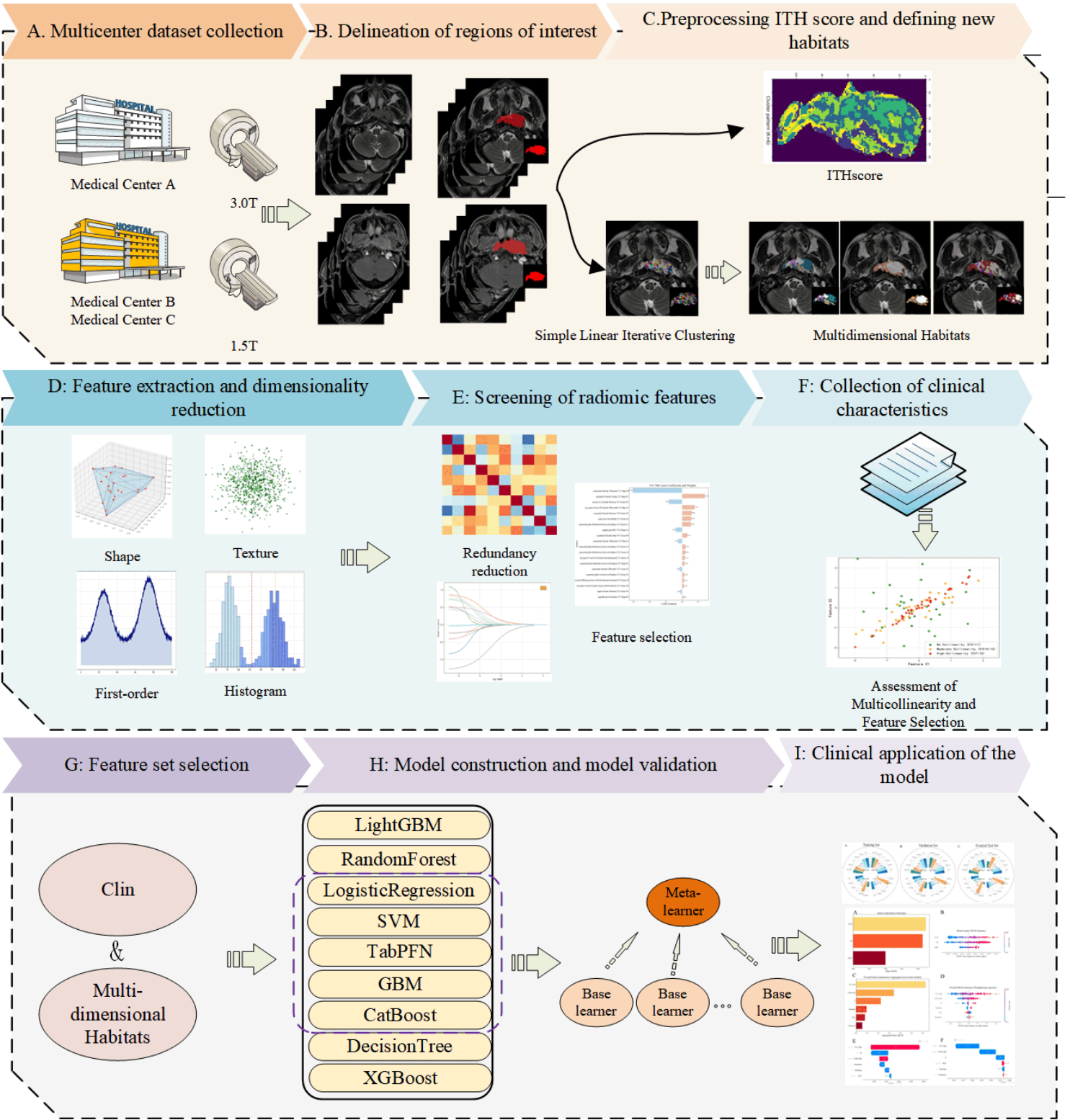


Figure 2

: Workflow of Multicenter Radiomics and Intratumoral Heterogeneity Analysis.

(A) MRI images (1.5T/3.0T) were collected from LANPC patients across three centers. ROIs were delineated on T2WI and T1C for feature extraction, including ITHscore, conventional radiomics, and Multidimensional Habitats.

(B) Clinical features were screened by VIF (<10) and refined via RFE, then integrated with the derived MH scores (T1C MH and T2WI MH) for modeling.

(C) Based on the integrated clinical and MH features, 10 base machine learning models were evaluated, from which the top 5 were selected to construct 26 stacking ensemble models. The final model's performance was validated on an external test set, and its clinical interpretability was demonstrated using SHAP visual analysis.

Abbreviations: MRI, magnetic resonance imaging; LANPC, locally advanced nasopharyngeal carcinoma; T2WI, T2-weighted imaging; T1C, contrast-enhanced T1-weighted imaging; VIF, variance inflation factor; ROI, region of interest; ITH, intratumoral heterogeneity; MH, multidimensional habitat; SHAP, Shapley Additive exPlanations.

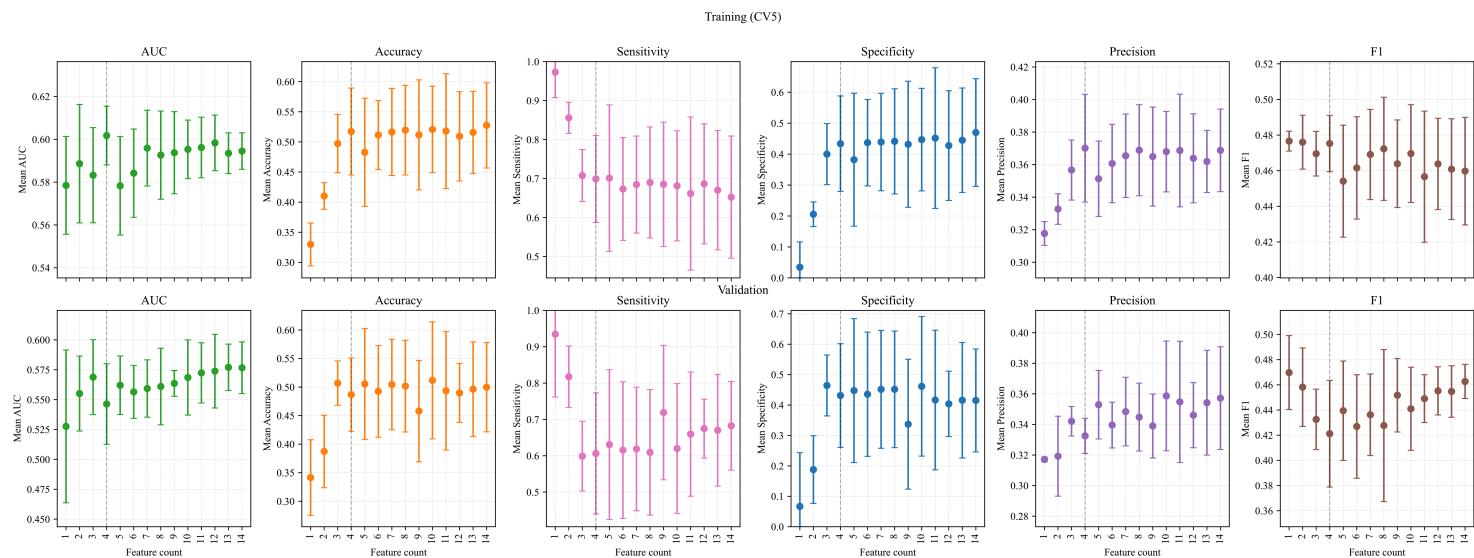


Figure 3

Impact of Different Feature Sets on Model Performance during Recursive Feature Elimination. The top row illustrates the performance metrics in the training set via 5-fold cross-validation (CV5), while the bottom row displays the corresponding performance in the independent validation set. From left to right, the metrics evaluated are AUC, accuracy, sensitivity, specificity, precision, and F1-score. All data points are presented as the mean $\pm$ 1 standard deviation.

Abbreviations: CV5, 5-fold cross-validation; AUC, area under the curve.

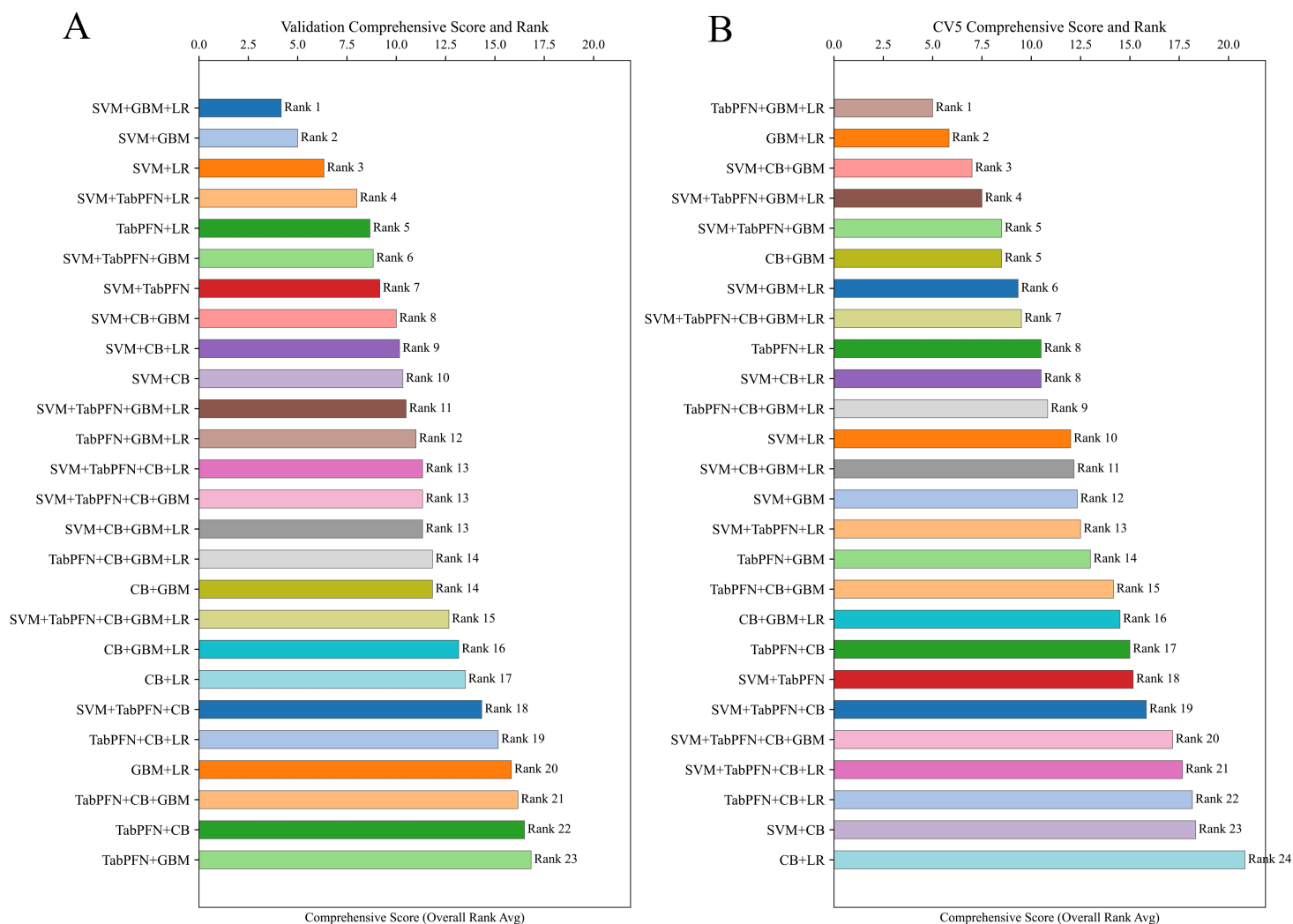


Figure 4

Comprehensive ranking of 26 stacking ensemble models constructed from the top five base learners. (A) Bar chart rankings based on performance in the validation set; (B) Bar chart rankings based on the 5-fold cross-validation (CV5) training set.

Abbreviations: CB, CatBoost; GBM, Gradient Boosting Machine; LR, Logistic Regression; SVM, Support Vector Machine; TabPFN, Tabular Prior-Data Fitted Network.

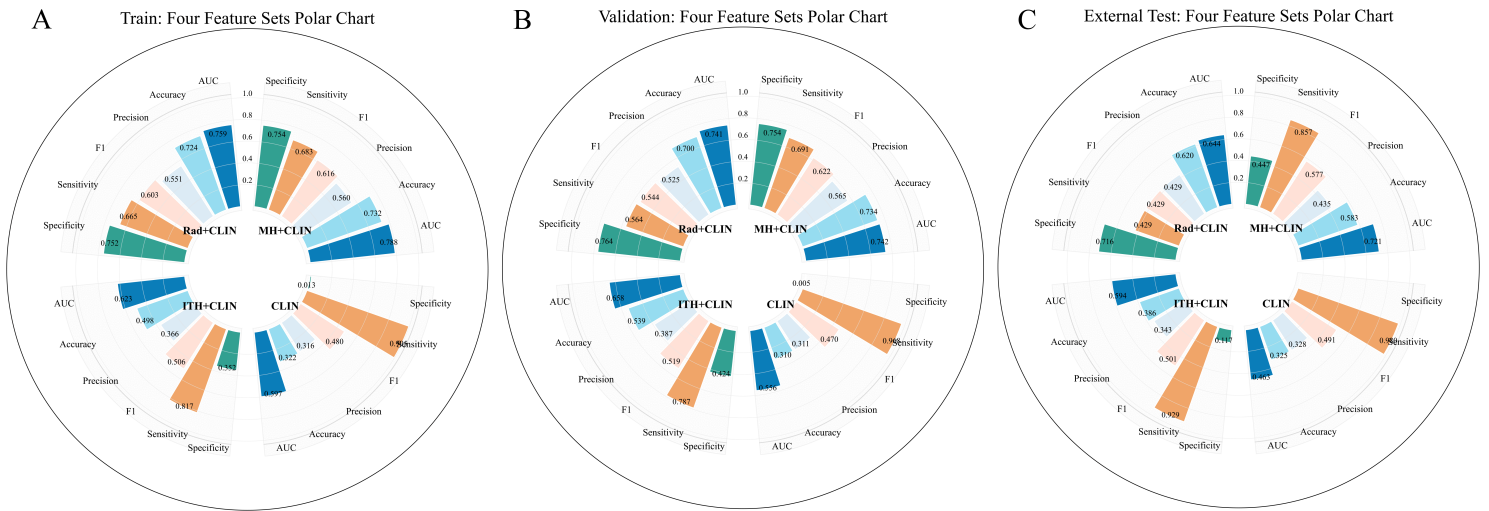


Figure 5

Polar coordinate visualization of performance metrics (AUC, Accuracy, Precision, F1-score, Sensitivity, and Specificity) for four feature sets: MH + Clin, Rad + Clin, ITH + Clin, and Clin. (a) Results on the validation set. (b) Results on the external test set.

Abbreviations: AUC, area under the receiver operating characteristic curve; Clin, clinical features; Rad, radiomics features; MH, multidimensional habitat; ITH, intratumoral heterogeneity.

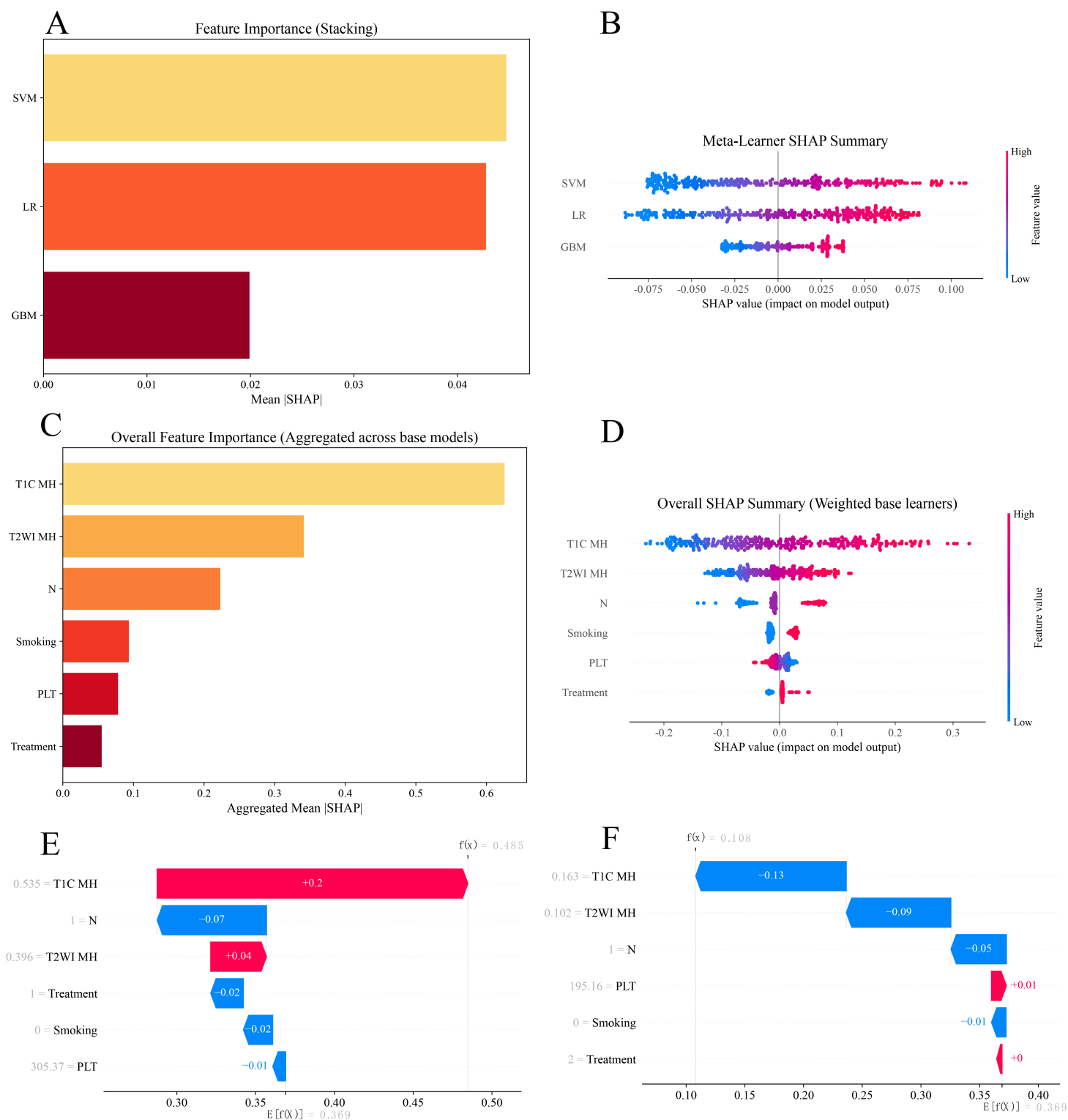


Figure 6

SHAP-based feature interpretation, visualization and case application analysis. (A) Importance ranking of base learners; (B) Distribution and direction of feature contributions across base learners; (C) Feature importance ranking of the meta-learner; (D) Distribution and direction of feature contributions in the meta-learner; (E) Local model interpretation for a high-risk individual, illustrating the role of each feature in the risk prediction for this case; (F) Local model interpretation for a low-risk individual.

Abbreviations: SHAP: SHapley Additive exPlanations; EBV: Epstein-Barr Virus; DNA: Deoxyribonucleic Acid; N: Lymph Node Classification.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [SupplementaryMultidimensionalHabitatstable1.docx](#)
- [Sfig5.tif](#)
- [Sfig1.tif](#)
- [Sfig8CALIBRATION.tif](#)
- [Sfig7DCA.tif](#)
- [Sfig6ROC.tif](#)
- [SupplementaryFigureslegends.docx](#)
- [Sfig3.tif](#)
- [Sfig4.tif](#)
- [Sfig2ROC.tif](#)