

Multi-modal Stacking Machine Learning Model for Predicting Lymph Node Metastasis in Cervical Cancer

Hong Yang

Ningxia Medical University

Wei Yang

yangwei_0521@163.com

General Hospital of Ningxia Medical University

Haiping Tian

General Hospital of Ningxia Medical University

Yadi Chang

Ningxia Medical University

Research Article

Keywords: Cervical cancer, Radiomics, Habitat radiomics, Machine learning, Stacking ensemble learning

Posted Date: May 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9636668/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

To develop and validate a multi-modal stacking machine learning model that integrates intratumoral and peritumoral habitat radiomics derived from multiparametric MRI with clinicopathological variables for non-invasive, preoperative prediction of lymph node metastasis (LNM) in early-stage cervical cancer (ECC).

Methods

This retrospective study with prospective validation analyzed 623 patients with ECC and divided them into training ($n = 311$), internal validation ($n = 187$), and external validation ($n = 125$) cohorts. Preoperative T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and delayed-phase contrast-enhanced T1-weighted imaging (CE-T1WI) were used to cluster tumor voxels into homogeneous habitats using the K-means algorithm based on signal intensity. Habitat radiomic features were extracted from intratumoral regions and a 3-mm peritumoral expansion. A stacking ensemble framework was constructed by combining five base classifiers (logistic regression, extreme gradient boosting, Elastic Net logistic regression, Depthwise Separable Convolutional Attention, and support vector machine) as first-level learners and an XGBoost-based Meta-model as the second-level learner. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), calibration curves, decision curve analysis, net reclassification improvement (NRI), and integrated discrimination improvement (IDI).

Results

Among the five base learners, the SVM-based ITH_integrated_score model achieved the best performance, with AUCs of 0.882, 0.865, and 0.845 in the training, internal validation, and external validation cohorts, respectively. The Meta-learner achieved AUCs of 0.915, 0.895, and 0.875, significantly outperforming all base learners (all $P < 0.05$). The Meta-model demonstrated excellent calibration and the highest clinical net benefit. NRI and IDI analyses confirmed significant incremental value over each base learner.

Conclusion

A multi-modal stacking ensemble machine learning model improves preoperative prediction of LNM compared with unimodal approaches and may help guide individualized surgical and surveillance strategies.

1. Background

Cervical cancer (CC) remains a substantial global health burden and is the fourth leading cause of cancer-related mortality among women worldwide [1]. For patients with early-stage cervical cancer (ECC), radical hysterectomy with pelvic lymph node dissection remains a standard treatment strategy [2]. However, fewer than 30% of patients with ECC have pelvic lymph node metastasis (LNM), indicating that more than 70% may undergo unnecessary lymphatic clearance and experience avoidable morbidity [3, 4]. Prognostically, five-year survival is reported to be 80%–100% in ECC without LNM, but decreases to 47%–78% when nodal metastasis is present [5]. Accurate preoperative identification of LNM is therefore essential to optimize treatment selection and reduce overtreatment.

Histopathological examination remains the definitive reference standard for LNM assessment, but it is inherently postoperative. Sentinel lymph node (SLN) biopsy has been advocated to reduce the extent of pelvic lymphadenectomy in ECC [6, 7]. The diagnostic yield of SLN evaluation may be improved by adjunct techniques such as immunohistochemistry, polymerase chain reaction, and serial sectioning (ultrastaging) [8, 9]. Nonetheless, both systematic lymphadenectomy and SLN biopsy are invasive procedures associated with surgical risks, false-negative results, and lymphatic or wound complications (e.g., lymphocele formation, lymphedema, neurovascular injury, and surgical site infection) [10, 11]. These limitations highlight the clinical need for reliable, noninvasive, preoperative tools that can estimate nodal status and support individualized decision-making.

Magnetic resonance imaging (MRI) is widely used for staging and treatment planning in cervical cancer [12]. Conventional MRI-based assessment of LNM largely depends on morphologic criteria such as nodal size and shape. Although MRI offers high specificity, its sensitivity for detecting LNM remains limited (reported sensitivity 27%–45.7%), particularly for normal-sized nodes and micrometastatic disease [13, 14].

Radiomics has emerged as a quantitative approach for outcome prediction in cervical cancer [15, 16]. However, most conventional radiomics workflows extract features from the entire tumor volume, which can average out spatially localized signal patterns and potentially dilute information associated with aggressive biology, treatment resistance, and adverse outcomes [17].

Habitat imaging seeks to address this limitation by partitioning tumors into heterogeneous subregions (“habitats”) based on imaging phenotypes that reflect underlying histopathologic and molecular variation. Each habitat may capture distinct biological attributes, including differences in angiogenesis, cellularity, metabolic activity, and gene expression, thereby preserving spatial heterogeneity for feature extraction [18, 19]. In parallel, tumor aggressiveness is influenced by alterations in the peritumoral microenvironment (e.g., stromal remodeling, angiogenic changes, and inflammatory infiltration), which can be characterized using peritumoral radiomics [20].

Integrating radiomics with clinicopathological and radiological variables presents methodological challenges because the relationships are often non-linear and high-dimensional. Traditional single-

model approaches, such as fixed-parameter logistic regression, may be insufficient to capture these complex interactions [21]. In contrast, machine learning methods can model non-linear associations and reduce overfitting through regularization and ensemble strategies. For example, Elastic Net regularization combines the strengths of Ridge and Lasso regression and is well suited to correlated, high-dimensional predictors [22]. Moreover, stacking ensembles, which combine multiple base learners through a meta-learner, can improve generalization by leveraging complementary model strengths [23, 24].

To our knowledge, no prior study has systematically developed a multi-modal stacking ensemble model that integrates clinicopathological features with both intratumoral and peritumoral habitat radiomics signatures for preoperative prediction of LNM in ECC. Therefore, we developed and validated a multi-modal stacking ensemble model to improve LNM prediction and support personalized treatment planning.

2. Materials and Methods

2.1. Patient Recruitment

This study protocol, including a retrospective analysis and a prospective validation phase, was approved by our institutional review board. Written informed consent was obtained from all patients enrolled in the prospective validation phase. From January 2020 to December 2025, 623 eligible patients were enrolled and allocated to a training cohort ($n = 311$), an internal validation cohort ($n = 187$), and an external validation cohort ($n = 125$). The inclusion criteria were as follows: (a) availability of complete multi-parametric MRI data acquired within two weeks before surgery; (b) undergoing radical hysterectomy with pelvic lymphadenectomy; and (c) postoperative histopathologic confirmation of cervical cancer with complete pelvic lymph node pathology results. The exclusion criteria were as follows: (a) prior treatment; (b) poor quality or incomplete MRI scans; (c) presence of other malignancies; (d) incomplete clinical data; and (e) absence of surgically confirmed nodal status. Figure 1 presents the inclusion, exclusion, and cohort allocation flowchart.

2.2. MRI Protocol

2.2.1. MR Imaging Acquisition and Image Preprocessing

The workflow of our research is shown in Fig. 2. All patients underwent pelvic MRI examinations before surgery using a 3.0-T MRI scanner (Signa Excite HD; GE Healthcare) equipped with an 8-channel phased-array coil. To reduce motion artifacts from bowel peristalsis, 20 mg of hyoscine butylbromide was administered intramuscularly after fasting for at least 6 h. Patients were positioned supine and scanned during free breathing. A half-filled urinary bladder was used as a reference to improve lesion conspicuity. The imaging protocol included T1-weighted imaging (T1WI), axial and sagittal fat-suppressed (FS) T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI) with b values of 0 and

1000 s/mm², and contrast-enhanced T1-weighted imaging (CE-T1WI). Detailed acquisition parameters are provided in Table S1. Morphological and functional MRI parameters of cervical cancer were extracted from conventional MRI sequences (Method S1)

Images were resampled to isotropic voxels of 1 mm³ using B-spline interpolation, and intensity values were standardized using histogram matching to improve comparability across scans [25].

2.2.2.ROI Segmentation and Habitat Generation

Three-dimensional tumor volumes of interest (VOIs) were manually segmented slice-by-slice on all sequences using 3D Slicer (version 5.0.3) by two board-certified radiologists (24 and 8 years of experience in gynecologic oncology imaging, respectively), both blinded to clinical outcomes and pathological data. The intratumoral VOI encompassed the entire tumor volume. The peritumoral VOI was defined as a 3-mm annular expansion beyond the tumor boundary. Segmentation reproducibility was assessed using intraclass correlation coefficients (ICCs), with an ICC threshold of ≥ 0.75 indicating acceptable agreement.

To address the challenge of visually delineating tumor subregions in ECC, we used a data-driven K-means clustering approach to generate intratumoral habitats [26]. The tumor volume was partitioned into spatially distinct subregions with consistent signal intensity patterns across DWI, T2WI, and CE-T1WI. Voxel intensities within the ROI across all patients were aggregated into a global matrix, with dimensions corresponding to the number of voxels and the three imaging modalities. Voxel-based K-means clustering was then performed to segment the target volume into heterogeneous habitats characterized by uniform intensity patterns. Cluster numbers ranging from 2 to 10 were evaluated, and the optimal cluster number was determined using the Calinski-Harabasz (CH) index [27]; the optimal number of clusters was 3 (Fig. S1).

2.3.Radiomic Feature Extraction, Selection, and Signature Construction

Habitat radiomic features were extracted from intratumoral habitats and peritumoral regions in accordance with the Image Biomarker Standardization Initiative (IBSI) guidelines [28]. Features were grouped into: (1) geometric metrics (shape and size), (2) first-order statistics (intensity distributions), and (3) texture features capturing spatial heterogeneity (GLCM, GLRLM, GLSZM, and NGTDM). Feature extraction was performed using PyRadiomics (version 3.0.1).

To ensure feature robustness and reduce dimensionality, a four-stage pipeline was implemented within the training cohort: (1) ICC-based reproducibility screening, (2) univariate significance testing (FDR-corrected $P < 0.05$), (3) redundancy elimination using Spearman correlation ($|\rho| > 0.90$), and (4) least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation (1-SE rule). This pipeline yielded 6 intratumoral features (1 from DWI, 2 from T2WI, and 3 from CE-T1WI), 9 peritumoral features (3 from each sequence), 10 combined intratumoral and peritumoral features (2

from DWI, 5 from T2WI, and 3 from CE-T1WI), and 6 intratumoral heterogeneity (ITH) features (2 from each sequence). The intratumoral radiomics score (Rad-score_intra), peritumoral radiomics score (Rad-score_peri), combined intratumoral–peritumoral radiomics score (Rad-score_intra+peri), and ITH_integrated_score were calculated as weighted linear combinations of the selected features using LASSO-derived coefficients. Detailed feature definitions, coefficients, and selection procedures are provided in Table S2 and Fig. S2.

2.4. Stacking Ensemble Machine Learning Model Construction

A two-tier stacking ensemble framework was developed to integrate prognostic information from clinical variables and habitat radiomics features. At the first tier, five base learners were constructed in parallel: a clinical logistic regression (LR) model incorporating independent prognostic factors identified in multivariable analysis, and four habitat radiomics models based on radiomic features (an XGBoost model trained on 6 intratumoral features, an Elastic Net LR model trained on 9 peritumoral features, a DSCA model trained on 10 combined features, and an SVM model trained on 6 ITH features to establish the ITH_integrated_score). Hyperparameters were optimized using 5-fold cross-validation to mitigate overfitting. At the second tier, the risk scores generated by the five base learners were combined using an XGBoost Meta-learner to produce the final ensemble prediction, enabling adaptive weighting of each constituent model.

2.5. Model Evaluation and Statistical Analysis

Model discrimination was assessed using receiver operating characteristic (ROC) curve analysis and the area under the curve (AUC). Calibration curves were used to evaluate agreement between predicted probabilities and observed outcomes. Clinical utility was assessed using decision curve analysis (DCA), integrated discrimination improvement (IDI), and category-free net reclassification index (NRI).

Statistical analyses were performed using SPSS (version 26.0), R (version 4.2.2), and Python (version 3.9.7). Continuous variables were tested for normality (Shapiro-Wilk test) and homoscedasticity (Levene's test). Normally distributed variables are presented as mean $\pm$ standard deviation and compared using independent-samples t tests. Non-normally distributed variables are presented as median (interquartile range, IQR) and compared using the Mann-Whitney U test. Categorical variables are presented as frequency (percentage) and compared using Pearson's χ^2 test or Fisher's exact test. Comparative model performance was assessed using the DeLong test. A two-tailed $P < 0.05$ was considered statistically significant.

3. Results

3.1. Patient Characteristics

Among 713 patients assessed for eligibility, 623 met the inclusion criteria and were included in the final analysis. The age range was 25–79 years (mean age $\pm$ standard deviation, 53.34 ± 12.34). The numbers of patients with LNM in the training, internal validation, and external validation cohorts were 74 (23.8%), 44 (23.5%), and 30 (24.0%), respectively. There were no significant differences in LNM positivity rates across the three cohorts ($P > 0.05$; Table 1), supporting the reliability of the external validation. In the training cohort, significant differences were observed between the LNM-positive and LNM-negative groups in age, tumor size, pathological type, pathological grade, parametrial infiltration, depth of stromal invasion, lymphovascular invasion (LVI), SCC antigen level, intratumoral Rad-score, peritumoral Rad-score, combined intratumoral–peritumoral Rad-score, and ITH_integrated_score (all $P < 0.05$; Table S3).

Table 1
Baseline clinicopathological characteristics of patients across the three cohorts.

Variable	Overall	Tra	Int-Val	Ext-Val	pa	pb	pc
LNM-positive rate	148 (23.8%)	74 (23.8%)	44 (23.5%)	30 (24.0%)	1.000	1.000	1.000
Age, mean $\pm$ SD	53.34 $\pm$ 12.34	53.89 $\pm$ 12.70	52.98 $\pm$ 11.66	52.48 $\pm$ 12.46	0.413	0.288	0.723
maximum_diameter, median (Q1, Q3)	27.18 (21.58, 34.23)	27.28 (21.52, 34.23)	26.29 (21.13, 33.34)	27.29 (22.22, 34.39)	0.440	0.574	0.253
Smoke history					0.235	0.506	0.849
No	442 (70.9%)	228 (73.3%)	127 (67.9%)	87 (69.6%)			
Yes	181 (29.1%)	83 (26.7%)	60 (32.1%)	38 (30.4%)			
HPV					0.329	0.446	0.114
Negative	86 (13.8%)	44 (14.1%)	20 (10.7%)	22 (17.6%)			
Positive	537 (86.2%)	267 (85.9%)	167 (89.3%)	103 (82.4%)			
Delivery times					0.442	0.273	0.093
< 3	348 (55.9%)	174 (55.9%)	112 (59.9%)	62 (49.6%)			
$\geq$ 3	275 (44.1%)	137 (44.1%)	75 (40.1%)	63 (50.4%)			
FIGO stage					0.179	0.262	1.000
Ⅰ	309 (49.6%)	144 (46.3%)	99 (52.9%)	66 (52.8%)			
Ⅱ	314 (50.4%)	167 (53.7%)	88 (47.1%)	59 (47.2%)			
Pathological_type					0.116	0.599	0.549
Squamous cell carcinoma	498 (79.9%)	256 (82.3%)	140 (74.9%)	102 (81.6%)			
Adenocarcinoma	82 (13.2%)	37 (11.9%)	31 (16.6%)	14 (11.2%)			
Adenosquamous carcinoma	28 (4.5%)	14 (4.5%)	9 (4.8%)	5 (4.0%)			

Variable	Overall	Tra	Int-Val	Ext-Val	pa	pb	pc
Other	15 (2.4%)	4 (1.3%)	7 (3.7%)	4 (3.2%)			
Pathological_grade					0.518	0.626	0.968
Moderate	103 (16.5%)	50 (16.1%)	31 (16.6%)	22 (17.6%)			
Poor	438 (70.3%)	225 (72.3%)	128 (68.4%)	85 (68.0%)			
Well	82 (13.2%)	36 (11.6%)	28 (15.0%)	18 (14.4%)			
Depth of stromal invasion					0.228	0.538	0.798
< 1/2	217 (34.8%)	116 (37.3%)	59 (31.6%)	42 (33.6%)			
≥ 1/2	406 (65.2%)	195 (62.7%)	128 (68.4%)	83 (66.4%)			
parametrial_infiltration					0.947	0.252	0.365
Negative	388 (62.3%)	189 (60.8%)	115 (61.5%)	84 (67.2%)			
Positive	235 (37.7%)	122 (39.2%)	72 (38.5%)	41 (32.8%)			
Ki67, median (Q1, Q3)	60.00 (52.50, 75.00)	60.00 (60.00, 80.00)	60.00 (50.00, 70.00)	60.00 (50.00, 75.00)	0.088	0.797	0.263
LVI					0.785	0.540	0.806
Negative	224 (36.0%)	108 (34.7%)	68 (36.4%)	48 (38.4%)			
Positive	399 (64.0%)	203 (65.3%)	119 (63.6%)	77 (61.6%)			
SCC-Ag, median (Q1, Q3)	1.90 (1.02, 7.04)	1.99 (1.02, 7.05)	1.90 (1.10, 7.04)	1.61 (1.00, 7.02)	0.929	0.461	0.404
ADC, median (Q1, Q3)	0.89 (0.73, 1.12)	0.89 (0.73, 1.12)	0.89 (0.69, 1.12)	0.90 (0.73, 1.12)	0.642	0.729	0.980

3.2. Variables Associated With LNM Positivity in the Training Cohort

Multivariable analysis demonstrated that tumor size (OR = 1.05, 95% CI: 1.02–1.08, $P < 0.001$), stromal invasion (OR = 1.82, 95% CI: 1.05–3.15, $P = 0.032$), parametrial infiltration (OR = 1.76, 95% CI: 1.01–3.08, $P = 0.045$), Rad-score_intra (OR = 4.22, 95% CI: 1.78–10.03, $P = 0.001$), Rad-score_peri (OR = 7.61, 95% CI: 2.14–27.06, $P < 0.001$), Rad-score_intra+peri

(OR = 15.18, 95% CI: 4.03–57.21, $P < 0.001$), and ITH_integrated_score

(OR = 24.37, 95% CI: 6.28–94.88, $P < 0.001$) were independent predictors of LNM positivity (Table 2; Fig. S3). The three clinicopathological variables (tumor size, depth of stromal invasion, and parametrial infiltration) were incorporated into the clinical LR base learner.

Table 2

Univariable and multivariable LR regression analyses of prognostic factors in the training cohort

Variable	Univariable analysis			Multivariable analysis		
	OR	95% CI	P value	OR	95% CI	P value
Age (years)	0.95	0.93–0.97	< 0.001	0.94	0.91–1.01	0.052
Tumour size (mm)	1.07	1.04–1.10	< 0.001	1.05	1.02–1.08	< 0.001
Smoking history (yes vs. no)	1.22	0.69–2.17	0.498			
HPV status (positive vs. negative)	0.93	0.44–1.94	0.839			
Delivery times (≥ 3 vs. < 3)	0.83	0.49–1.41	0.486			
FIGO stage	1.37	0.91–2.05	0.130			
Pathological type (Non-squamous vs. Squamous)	0.27	0.10–0.71	0.008			
Pathological grade (per grade increase)	0.45	0.27–0.76	0.003			
Depth of stromal invasion ($\geq 1/2$ vs. $< 1/2$)	2.41	1.32–4.38	0.004	1.82	1.05–3.15	0.032
Parametrial infiltration (positive vs. negative)	2.22	1.31–3.78	0.003	1.76	1.01–3.08	0.045
Ki-67 index (%)	0.99	0.98–1.00	0.147			
LVI (positive vs. negative)	1.95	1.12–3.41	0.018	1.65	0.92–2.95	0.092
SCC-Ag (ng/mL)	1.02	1.01–1.04	0.004	1.01	1.00–1.03	0.097
ADC ($\times 10^{-3}$ mm ² /s)	0.57	0.23–1.42	0.227	0.59	0.22–1.60	0.304
Rad-score_intra	4.22	1.78–10.03	0.001	2.15	1.10–4.20	0.024

Abbreviations: Variables with $p < 0.05$ in the univariable analysis were entered into the multivariable model. OR, odds ratio; CI confidence interval; Rad-score_intra, intratumoral radiomics score; Rad-score_peri, peritumoral radiomics score; Rad-score_intra+peri, combined intratumoral and peritumoral radiomics score; ITH_integrated_score Intratumoral heterogeneity score integrated intratumoral and peritumoral.

Variable	Univariable analysis			Multivariable analysis		
	OR	95% CI	P value	OR	95% CI	P value
Rad-score_peri	7.61	2.14–27.06	< 0.001	3.20	1.45–7.05	0.004
Rad-score_intra+peri	15.18	4.03–57.21	< 0.001	12.36	3.52–43.45	< 0.001
ITH_integrated_score	24.37	6.28–94.88	< 0.001	18.45	4.88–69.76	< 0.001
Abbreviations: Variables with $p < 0.05$ in the univariable analysis were entered into the multivariable model. OR, odds ratio; CI confidence interval; Rad-score_intra, intratumoral radiomics score; Rad-score_peri, peritumoral radiomics score; Rad-score_intra+peri, combined intratumoral and peritumoral radiomics score; ITH_integrated_score Intratumoral heterogeneity score integrated intratumoral and peritumoral.						

3.3. Performance Evaluation of Prediction Models

Overall model discrimination and classification performance across cohorts are summarized in Table 3 and Fig. 3. The Meta-model achieved AUC values of 0.915, 0.895, and 0.875 in the training, internal validation, and external validation cohorts, respectively, and outperformed each comparator model based on DeLong testing (all $P < 0.05$; Table S4). Calibration curves demonstrated good agreement between predicted probabilities and observed outcomes (Fig. 3c). Decision curve analysis indicated that the stacking model provided higher net benefit across a broad range of threshold probabilities (Fig. 3d). Incremental performance gains were further supported by NRI and IDI analyses (Table 4; Fig. 4).

Table 3
Diagnostic performance of the base models and meta-model across the three cohorts

Cohort	Metric	LR	XGBoost	Elastic Net LR	DSCA	SVM	Meta
Tra	AUC (95% CI)	0.685 (0.610– 0.740)	0.805 (0.750– 0.860)	0.752 (0.690– 0.810)	0.854 (0.800– 0.900)	0.882 (0.830– 0.930)	0.915 (0.880– 0.950)
Tra	Accuracy (95% CI)	0.640 (0.585– 0.695)	0.750 (0.700– 0.800)	0.695 (0.640– 0.750)	0.795 (0.745– 0.840)	0.820 (0.775– 0.865)	0.860 (0.815– 0.895)
Tra	Sensitivity (95% CI)	0.650 (0.535– 0.755)	0.760 (0.655– 0.845)	0.705 (0.595– 0.800)	0.810 (0.710– 0.885)	0.840 (0.745– 0.910)	0.880 (0.790– 0.940)
Tra	Specificity (95% CI)	0.635 (0.570– 0.695)	0.745 (0.685– 0.800)	0.685 (0.620– 0.745)	0.785 (0.725– 0.835)	0.810 (0.755– 0.860)	0.850 (0.800– 0.895)
Tra	PPV (95% CI)	0.355 (0.280– 0.435)	0.485 (0.395– 0.575)	0.410 (0.325– 0.495)	0.540 (0.445– 0.630)	0.580 (0.485– 0.670)	0.650 (0.555– 0.735)
Tra	NPV (95% CI)	0.855 (0.800– 0.900)	0.910 (0.865– 0.945)	0.880 (0.830– 0.920)	0.930 (0.885– 0.960)	0.945 (0.905– 0.970)	0.960 (0.925– 0.985)
Tra	F1-score (95% CI)	0.455 (0.380– 0.530)	0.590 (0.510– 0.670)	0.515 (0.435– 0.595)	0.645 (0.565– 0.720)	0.685 (0.605– 0.760)	0.750 (0.675– 0.815)
Int-Val	AUC (95% CI)	0.665 (0.585– 0.735)	0.785 (0.715– 0.840)	0.735 (0.665– 0.795)	0.835 (0.775– 0.885)	0.865 (0.805– 0.910)	0.895 (0.845– 0.935)
Int-Val	Accuracy (95% CI)	0.625 (0.550– 0.695)	0.735 (0.665– 0.795)	0.680 (0.610– 0.745)	0.775 (0.710– 0.830)	0.805 (0.740– 0.860)	0.840 (0.780– 0.890)
Int-Val	Sensitivity (95% CI)	0.615 (0.465– 0.750)	0.735 (0.585– 0.855)	0.685 (0.535– 0.810)	0.785 (0.640– 0.895)	0.815 (0.675– 0.915)	0.855 (0.720– 0.945)
Int-Val	Specificity (95% CI)	0.625 (0.540– 0.705)	0.735 (0.655– 0.805)	0.675 (0.595– 0.750)	0.770 (0.695– 0.835)	0.800 (0.725– 0.865)	0.835 (0.765– 0.890)
Int-Val	PPV (95% CI)	0.335 (0.245– 0.440)	0.460 (0.350– 0.575)	0.395(0.295– 0.505)	0.510 (0.395– 0.625)	0.555 (0.435– 0.670)	0.615 (0.490– 0.730)
Int-Val	NPV (95% CI)	0.840 (0.765– 0.895)	0.895 (0.830– 0.940)	0.870 (0.800– 0.920)	0.915 (0.855– 0.955)	0.930 (0.875– 0.965)	0.945 (0.895– 0.975)

Cohort	Metric	LR	XGBoost	Elastic Net LR	DSCA	SVM	Meta
Int-Val	F1-score (95% CI)	0.430 (0.335–0.535)	0.565 (0.460–0.670)	0.495 (0.395–0.600)	0.615 (0.510–0.720)	0.660 (0.555–0.760)	0.715 (0.615–0.810)
Ext-Val	AUC (95% CI)	0.650 (0.545–0.745)	0.765 (0.665–0.845)	0.715 (0.615–0.800)	0.815 (0.720–0.885)	0.845 (0.755–0.910)	0.875 (0.795–0.935)
Ext-Val	Accuracy (95% CI)	0.610 (0.520–0.695)	0.715 (0.630–0.790)	0.665 (0.575–0.745)	0.755 (0.675–0.825)	0.785 (0.705–0.850)	0.820 (0.745–0.880)
Ext-Val	Sensitivity (95% CI)	0.590 (0.405–0.760)	0.700 (0.515–0.845)	0.650 (0.460–0.810)	0.750 (0.565–0.885)	0.785 (0.605–0.910)	0.820 (0.645–0.935)
Ext-Val	Specificity (95% CI)	0.615 (0.515–0.710)	0.715 (0.615–0.800)	0.665 (0.565–0.755)	0.755 (0.660–0.835)	0.785 (0.690–0.860)	0.820 (0.730–0.890)
Ext-Val	PPV (95% CI)	0.320 (0.210–0.450)	0.440 (0.310–0.575)	0.380 (0.260–0.515)	0.490 (0.355–0.625)	0.535 (0.395–0.670)	0.590 (0.445–0.725)
Ext-Val	NPV (95% CI)	0.825 (0.730–0.895)	0.880 (0.795–0.940)	0.855 (0.765–0.920)	0.905 (0.825–0.955)	0.920 (0.845–0.965)	0.935 (0.865–0.975)
Ext-Val	F1-score (95% CI)	0.415 (0.295–0.540)	0.540 (0.415–0.665)	0.475 (0.355–0.605)	0.595 (0.465–0.715)	0.635 (0.505–0.755)	0.685 (0.555–0.800)

Table 4

Incremental predictive value of the Meta-Model over other stacking models assessed by NRI and IDI.

Model Comparison	Tra NRI (95% CI)	Tra IDI (95% CI)	Int NRI (95% CI)	Int IDI (95% CI)	Ext NRI (95% CI)	Ext IDI (95% CI)
Stacking1 [P1, P2, P3, P4] (vs Meta-Model)	0.385 (0.210–0.480)	0.255 (0.150–0.350)	0.312 (0.180–0.420)	0.198 (0.110–0.290)	0.285 (0.150–0.410)	0.165 (0.090–0.250)
Stacking2 [P1, P2, P3, P4, ITH_integrated_score] (vs Meta-Model)	0.185 (0.050–0.320)	0.085 (0.020–0.150)	0.152 (0.040–0.280)	0.068 (0.015–0.120)	0.125 (0.020–0.250)	0.055 (0.010–0.110)
Abbreviations: Tra, training cohort; Int, internal validation cohort; Ext, external validation cohort; NRI, net reclassification improvement; IDI, integrated discrimination improvement; CI, confidence interval (all values are presented as estimate with 95% CI).						

3.4. Interpretability Analysis

To elucidate the decision-making process of the stacking model, Shapley Additive Explanations (SHAP) analysis was performed. The global SHAP summary plot indicated that the SVM-based ITH_integrated_score model exhibited the highest mean absolute SHAP value, followed by the DSCA, XGBoost, Elastic Net LR, and clinical LR models. These findings suggest that the integrated ITH-score model and the combined intratumoral–peritumoral habitat radiomic representation were the dominant drivers of prediction (Fig. 5). SHAP-based visualizations for two representative cases are provided in Fig. 6. Positive SHAP values from multiple modalities increased the predicted probability of LNM, whereas negative SHAP values reduced the predicted probability.

4..Discussion

In this study, we developed and validated a multi-modal stacking machine learning (ML) model for noninvasive, preoperative prediction of LNM in ECC. By integrating clinicopathological variables with intratumoral habitat radiomics, peritumoral radiomics, and the ITH_integrated_score model, the proposed XGBoost- based stacking framework demonstrated strong discriminative performance. In the training cohort, the stacking model achieved an AUC of 0.915 (95% CI: 0.880–0.950) and significantly outperformed single-modality models (De- Long test, $P < 0.05$). This performance was confirmed in the external validation cohort (AUC = 0.875, 95% CI: 0.795–0.935). The stacking model also yielded greater and more consistent net benefit on decision curve analysis (DCA) and showed good calibration. Net reclassification improvement (NRI) and integrated discrimination improvement (IDI) analyses indicated incremental value over each base learner. SHAP-based interpretability analyses further suggested that the ITH_integrated_score and the combined intratumoral–peritumoral habitat radiomics representation contributed substantially to model output, exceeding the contribution of the clinical model. Collectively, these findings support stacking ensembles as a practical framework for integrating complementary information across modalities and improving preoperative LNM prediction in ECC.

Among unimodal models, the ITH_integrated_score model and the combined intratumoral–peritumoral habitat radiomics model showed higher performance in the external validation cohort (AUC = 0.845, 95% CI: 0.755– 0.910; and AUC = 0.815, 95% CI: 0.720–0.885, respectively), outperforming the intratumoral-only and peritumoral-only models (AUCs = 0.765 and 0.715). This pattern is consistent with prior efforts to predict LNM in ECC using MRI habitat radiomics [29]. Methodologically, habitat-based radiomics aims to preserve intratumoral spatial heterogeneity by partitioning tumors into imaging-defined subregions, thereby complementing conventional whole- tumor radiomics and providing a more biologically informed representation of intratumoral heterogeneity [17–19]. In addition, habitat analysis has been explored in other oncologic settings to characterize heterogeneous tumor phenotypes and relate them to clinically relevant endpoints [31, 32]. In our analysis, the most predictive habitat features were derived from the H1 subregion, corresponding to low-density areas on imaging. Pathologically, such regions are often associated with intratumoral hypoxia and necrosis, which are microenvironmental factors that can promote invasion, angiogenesis, and aberrant cellular proliferation. These observations support the potential clinical translatability of habitat radiomics for biologically grounded preoperative LNM risk stratification in ECC.

Another key finding is that peritumoral radiomics (Rad-score_peri) improved predictive performance. The tumor microenvironment is increasingly recognized as a driver of malignant behavior, including invasion, metastasis, immune evasion, and angiogenesis [33, 37]. Prior studies have also reported the value of peritumoral radiomics for predicting parametrial invasion and for prognostic assessment [34, 35]. In our study, peritumoral habitat analysis provides a complementary approach to capturing tumor microenvironment interactions, and the performance of the 3-mm peritumoral expansion suggests a reasonable spatial scale that may capture invasion-related changes while limiting dilution from more distant tissues. From a pathophysiological perspective, intratumoral features likely reflect cellularity, necrosis, and heterogeneity—surrogates of aggressiveness and treatment resistance— whereas peritumoral features may encode lymphovascular space invasion, stromal desmoplasia, and inflammatory infiltration, which are recognized drivers of recurrence and metastasis [36, 37].

Beyond habitat radiomics, an additional methodological contribution of this work is the integration of habitat radiomics with stacking ensemble learning. The rationale is that heterogeneous, multi-scale prognostic information may not be fully captured by any single modeling paradigm. Logistic regression offers interpretability but is constrained by linearity assumptions, while more flexible approaches can model non-linearities at the cost of increased overfitting risk [22]. Stacking ensembles can improve performance in radiomics by aggregating complementary predictions from multiple learners via a meta-learner [39, 40]. To support robust training of the meta-learner and reduce data leakage, we adopted an out-of-fold strategy within a 5-fold cross-validation framework. Regularization measures (restricted tree depth, increased minimum leaf sample requirements, and an L2 penalty) were applied to mitigate overfitting and improve generalization.

We further used SHAP to enhance interpretability by providing global and local explanations of how model components contribute to predicted LNM probability. SHAP highlighted the SVM-based ITH_integrated_score model as a key contributor, reinforcing the importance of intratumoral heterogeneity for improving prediction accuracy and supporting potential clinical adoption through increased transparency.

Clinical multivariable analysis identified tumor size, deep stromal invasion, and parametrial infiltration as independent predictors of LNM. Tumor size is a surrogate of tumor burden, while deep stromal invasion and parametrial infiltration reflect local and regional spread. The observed association patterns may vary across studies, potentially due to differences in cohort composition and class imbalance.

This study has limitations. First, although external validation was performed, the study is not fully multi-center; further validation in larger multi-institutional cohorts is warranted. Second, manual tumor delineation may affect radiomic feature stability. Although we performed reproducibility screening using ICCs (> 0.75), automated segmentation could improve efficiency and reproducibility. Third, due to limited long-term follow-up, we could not comprehensively evaluate postoperative recurrence; future studies with longer follow-up are needed to assess long-term outcomes.

5. Conclusion

We developed and validated a multi-modal stacking machine learning model that integrates clinicopathological variables with intratumoral and peritumoral habitat radiomic features for prediction of in ECC. The meta-model demonstrated superior discriminative performance across cohorts and may provide a clinically actionable framework to support surgical decision-making, risk- adapted surveillance, and treatment intensification.

Abbreviations

All
overall cohort (n = 623)
Trn
training cohort (n = 311)
Int-Val
internal validation cohort (n = 187)
Ext-Val
external validation cohort (n = 125)
SD
standard deviation
Q1
first quartile
Q3
third quartile
HPV
human papillomavirus
FIGO
International Federation of Gynecology and Obstetrics
ca.
carcinoma
Ki-67
Ki67 proliferation index
LVI
lymphovascular invasion
LN
lymph node
SCC-Ag
squamous cell carcinoma antigen
ADCmin
minimum apparent diffusion coefficient. pa, comparison between Trn and Int-Val
pb

comparison between Trn and Ext-Val
pc
comparison between Int-Val and Ext-Val.

Declarations

Author information

Hong Yang ^a, Wei Yang ^{b,c,*}, Haiping Tian ^d, Yadi Chang ^a

^a College of Clinical Medicine, Ningxia Medical University, 692 Shengli Road, Yinchuan 750004, China

^b Department of Radiology, General Hospital of Ningxia Medical University, 804 Shengli Road, Yinchuan 750004, China

^c State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Xinjiang Medical University, Wulumuqi 830011, China

^d Department of Pathology, General Hospital of Ningxia Medical University, 804 Shengli Road, Yinchuan 750004, China

* **Corresponding author** : Wei Yang, E-mail: yangwei_0521@163.com

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to patient privacy regulations but are available from the corresponding author on reasonable request, subject to institutional data sharing agreements and ethics committee approval.

Author contribution

Hong Yang: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Wei Yang: Writing – review & editing, Formal analysis, Conceptualization, Resources, Funding acquisition.

Haiping Tian: Resources, Methodology, Conceptualization.

Yadi Chang: Investigation, Data curation, Validation.

Acknowledgements

None.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the General Hospital of Ningxia Medical University (Protocol No. KYLL-2018-062) and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all prospectively enrolled participants. The requirement for individual consent was waived for retrospectively enrolled patients by the ethics committee due to the observational nature of the study and use of de-identified data.

Consent for publication

Not applicable. No individually identifiable data or images are presented.

Competing interests

The authors declare no competing interests.

Funding

This study was supported by the State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia Fund(SKL-HIDCA-2024-NX4) and National Natural Science Foundation of China[81860302].

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74:229-263.
2. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet.* 2009;105:103-104.
3. Bats A, Frati A, Mathevet P. Contribution of lymphoscintigraphy to intraoperative sentinel lymph node detection in early cervical cancer: Analysis of the prospective multicenter SENTICOL cohort. *Gynecol Oncol.* 2015;137:264-269.
4. Achouri A, Huchon C, Bats A, Bensaid C, Nos C, Lecuru F. Complications of lymphadenectomy for gynecologic cancer. *Eur J Surg Oncol.* 2013;39:81-86.
5. Gien LT, Covens A. Lymph node assessment in cervical cancer: prognostic and therapeutic implications. *J Surg Oncol.* 2009;99:242-247.
6. Cormier B, Diaz JP, Shih K. Establishing a sentinel lymph node mapping algorithm for the treatment of early cervical cancer. *Gynecol Oncol.* 2011;122:275-280.
7. Lecuru F, Mathevet P, Querleu D. Bilateral negative sentinel nodes accurately predict absence of lymph node metastasis in early cervical cancer: results of the SENTICOL study. *J Clin Oncol.* 2011;29:1686-1691.

8. Bats A, Mathevet P, Buenerd A. The sentinel node technique detects unexpected drainage pathways and allows nodal ultrastaging in early cervical cancer: insights from the multicenter prospective SENTICOL study. *Ann Surg Oncol*. 2013;20:413-422.
9. Kinkel K. Pitfalls in staging uterine neoplasm with imaging: a review. *Abdom Imaging*. 2006;31:164-173.
10. Zikan M, Fischerova D, Pinkavova I, Slama J, Weinberger V, Dusek L. A prospective study examining the incidence of asymptomatic and symptomatic lymphoceles following lymphadenectomy in patients with gynecological cancer. *Gynecol Oncol*. 2015;137:291-298.
11. Hwang JH, Kim BW. The incidence of perioperative lymphatic complications after radical hysterectomy and pelvic lymphadenectomy between robotic and laparoscopic approach: a systemic review and meta-analysis. *Int J Surg*. 2023;109:2478-2485.
12. Zhang W, Chen C, Liu P, Li W, Hao M, Zhao W. Impact of pelvic MRI in routine clinical practice on staging of IB1–IIA2 cervical cancer. *Cancer Manag Res*. 2019;11:3603-3609.
13. Woo S, Suh CH, Kim SY, Cho JY, Kim SH. The diagnostic performance of MRI for detection of lymph node metastasis in bladder and prostate cancer: an updated systematic review and diagnostic meta-analysis. *AJR Am J Roentgenol*. 2018;210:W95-W109.
14. Scheidler J, Hricak H, Yu KK, Subak L, Segal MR. Radiological evaluation of lymph node metastases in patients with cervical cancer: A meta-analysis. *JAMA*. 1997;278(13):1096-1101.
15. Sun C, Tian X, Liu Z. Radiomic analysis for pretreatment prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer: a multicentre study. *EBioMedicine*. 2020;46:160-169.
16. Fang J, Zhang B, Wang S. Association of MRI-derived radiomic biomarker with disease-free survival in patients with early-stage cervical cancer. *Theranostics*. 2020;10:2284-2292.
17. Andor N, Graham TA, Jansen M, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med*. 2016;22:105-113.
18. Wang S, Liu X, Wu Y, 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.
19. Wu J, Cao G, Sun X, 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.
20. Xiao ML, Fu L, Wei Y. Intratumoral and peritumoral MRI radiomics nomogram for predicting parametrial invasion in patients with early-stage cervical adenocarcinoma and adenosquamous carcinoma. *Eur Radiol*. 2024;34:852-862.
21. Wu G, Woodruff HC, Shen J. Diagnosis of invasive lung adenocarcinoma based on chest CT radiomic features of part-solid pulmonary nodules: A multicenter study. *Radiology*. 2020;297:E282.
22. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. *ACM Comput Surv*. 2019;51:1-36.

23. Zeng L, Ma P, Li Z. Multimodal Machine Learning-Based Marker Enables Early Detection and Prognosis Prediction for Hyperuricemia. *Adv Sci.* 2024;11:e2404047.
24. Binson VA, Subramoniam M, Ragesh GK, Kumar A. Early detection of lung cancer through breath analysis using adaboost ensemble learning method. 2021 2nd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS). 2021;183-187.
25. Nyul LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging.* 2000;19:143-150.
26. Timmerman ME, Ceulemans E, De Roover K, Van Leeuwen K. Sub- space K-means clustering. *Behav Res Methods.* 2013;45:1011-1023.
27. Zhang W, Yue Z, Ye J. Modulation format identification using the Calinski-Harabasz index. *Appl Opt.* 2022;61:851-857.
28. Zwanenburg A, Vallières M, Abdalah MA. The image biomarker standardization initiative: Standardized quantitative radiomics for high- throughput image-based phenotyping. *Radiology.* 2020;295:328-338.
29. Wang M, Cao Y, Zhang W, Liang Y, Liu J, Lei J. Prediction Model of Lymph Node Metastasis in Cervical Cancer Based on MRI Habitat Radiomics. *Cancers (Basel).* 2025;18:152.
30. Wang X, Su R, Li L. Machine learning-based radiomics for predicting outcomes in cervical cancer patients undergoing concurrent chemoradiotherapy. *Comput Biol Med.* 2024;177:108593.
31. Lee DH, Park JE, Kim N. Tumor habitat analysis by magnetic resonance imaging distinguishes tumor progression from radiation necrosis in brain metastases after stereotactic radiosurgery. *Eur Radiol.* 2022;32:497-507.
32. Stringfield O, Arrington JA, Johnston SK. Multiparameter MRI Predictors of Long-Term Survival in Glioblastoma Multiforme. *Tomography.* 2019;5:135-144.
33. Bi K, He MX, Bakouny Z. Tumor and immune reprogramming during immunotherapy in advanced renal cell carcinoma. *Cancer Cell.* 2025;43:1177-1179.
34. Xiao ML, Fu L, Wei Y. Intratumoral and peritumoral MRI radiomics nomogram for predicting parametrial invasion in patients with early- stage cervical adenocarcinoma and adenosquamous carcinoma. *Eur Radiol.* 2024;34:852-862.
35. Zhang Y, Wu C, Du J. Prediction of recurrence risk factors in patients with early-stage cervical cancers by nomogram based on MRI handcrafted radiomics features and deep learning features: a dual-center study. *Abdom Radiol.* 2024;49:258-270.
36. Cui L, Yu T, Kan Y. Multi-parametric MRI-based peritumoral radiomics on prediction of lymphovascular space invasion in early- stage cervical cancer. *Diagn Interv Radiol.* 2022;28:312-321.
37. Khorrami M, Prasanna P, Gupta A. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer. *Cancer Immunol Res.* 2020;8:108-119.

38. Wang P, Li Y, Reddy CK. Machine learning for survival analysis: a survey. ACM Comput Surv. 2019;51:1-36.

39. Fu Y, Wang X, Yi X, Guan X, Chen C, Han Z. Ensemble machine learning model incorporating radiomics and body composition for predicting intraoperative HDI in PPGL. J Clin Endocrinol Metab. 2024;109:351-360.

40. Gong J, Wang Ti, Wang Z, Chu X, Hu T, Li M. Enhancing brain metastasis prediction in non-small cell lung cancer: a deep learning-based segmentation and CT radiomics-based ensemble learning model. Cancer Imaging. 2024;24:1.

Figures

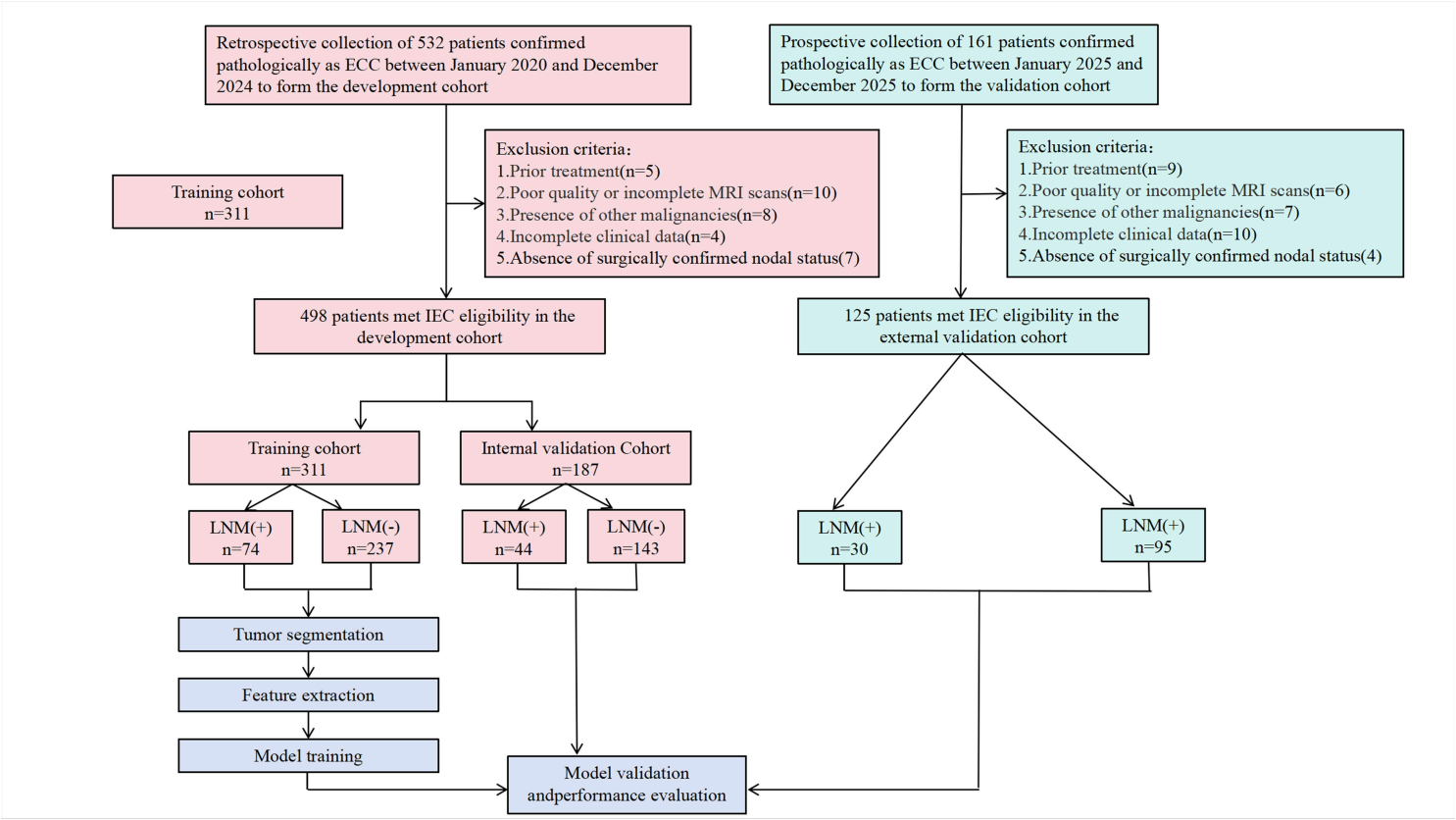


Figure 1

Patient enrollment flowchart and study design schema. The study comprised a retrospective training cohort (n = 311), an internal validation cohort (n = 187), and a prospectively collected external validation cohort (n = 125).

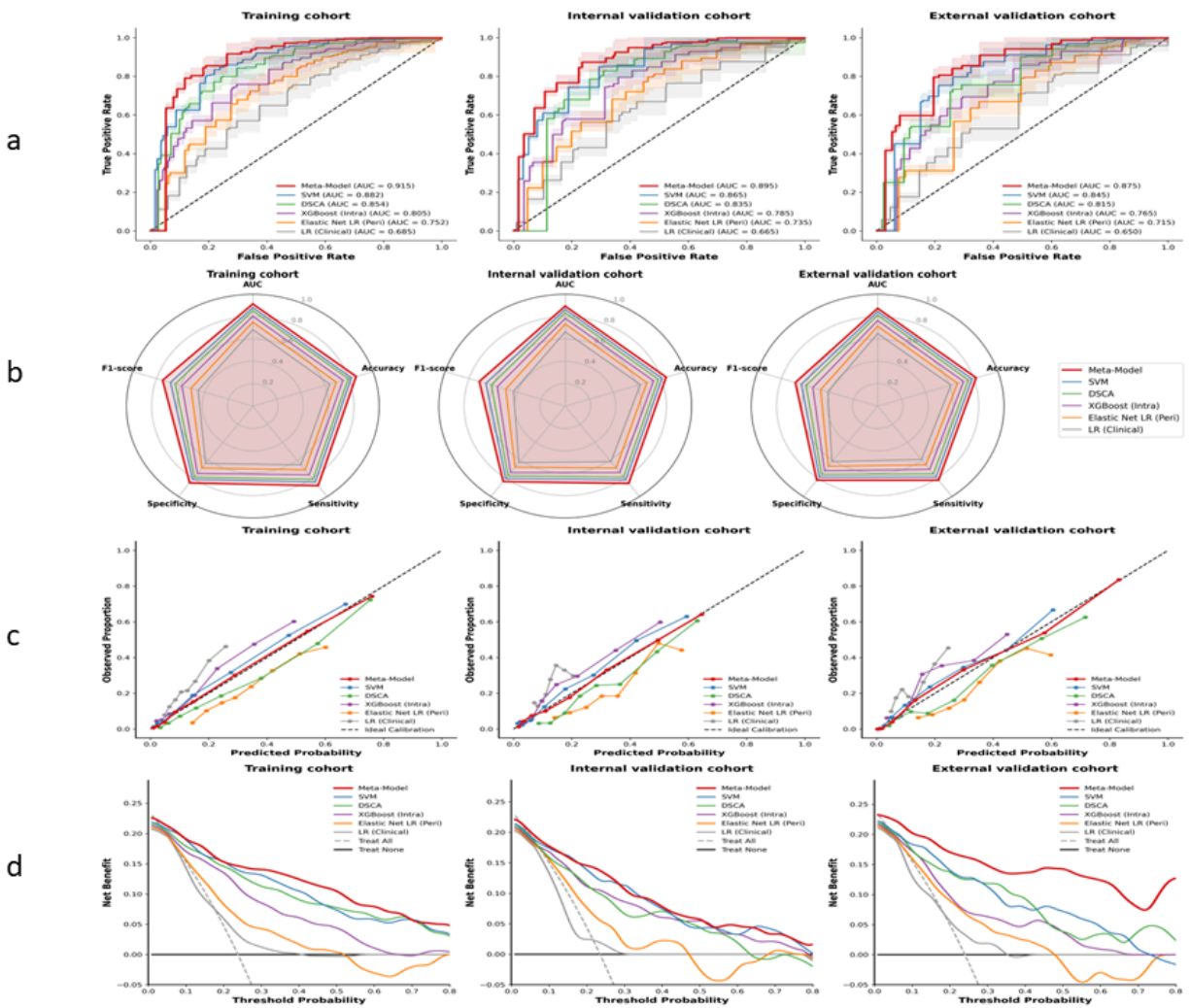


Figure 3

Comparative performance evaluation of the Meta-Model and individual base learners.

(a) Receiver operating characteristic (ROC) curves in the training cohort, internal validation cohort, and external validation cohort. The Meta-Model (red line) achieves the highest area under the curve (AUC) values across all three cohorts (0.915, 0.895, and 0.875, respectively). (b) Radar charts comparing model performance across seven key metrics (Sensitivity, Specificity, Accuracy, AUC, PPV, NPV, and F1-Score) in the training cohort, internal validation cohort, and external validation cohort. The Meta-Model consistently encompasses the largest area, demonstrating the most robust and balanced performance across all evaluation dimensions. (c) Calibration curves in the training cohort, internal validation cohort, and external validation cohort. The x-axis represents the predicted probability, and the y-axis represents the actual probability, with the dashed 45-degree line indicating perfect calibration. The Meta-Model demonstrates superior predictive accuracy and consistency with the actual outcomes across all three cohorts, as indicated by its closer proximity to the ideal diagonal line and lower Brier scores. (d) Decision curve analysis (DCA) in the training cohort, internal validation cohort, and external validation cohort. The Meta-Model demonstrates the highest net benefit across the clinically relevant threshold probability range (0.20–0.50).

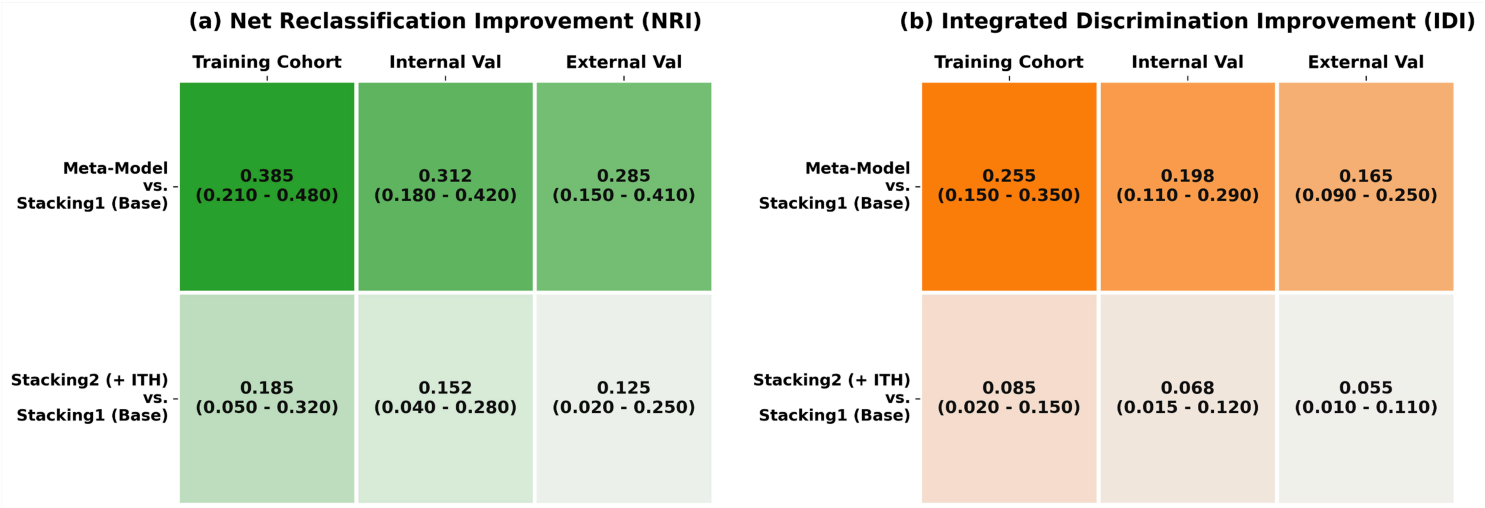


Figure 4

Incremental predictive value of the Meta-Model over other stacking models assessed by NRI and IDI. Heatmaps of Net Reclassification Index (NRI) and Integrated Discrimination Improvement (IDI) comparing the Meta-Model with other stacking models in the training cohort, internal validation cohort (c, d), and external validation cohort. The color intensity represents the magnitude of the improvement value, with warmer colors indicating positive improvement and cooler colors indicating negative improvement or decline in performance when comparing the new model against the reference baseline. The Meta-Model consistently demonstrates superior reclassification performance across all cohorts.

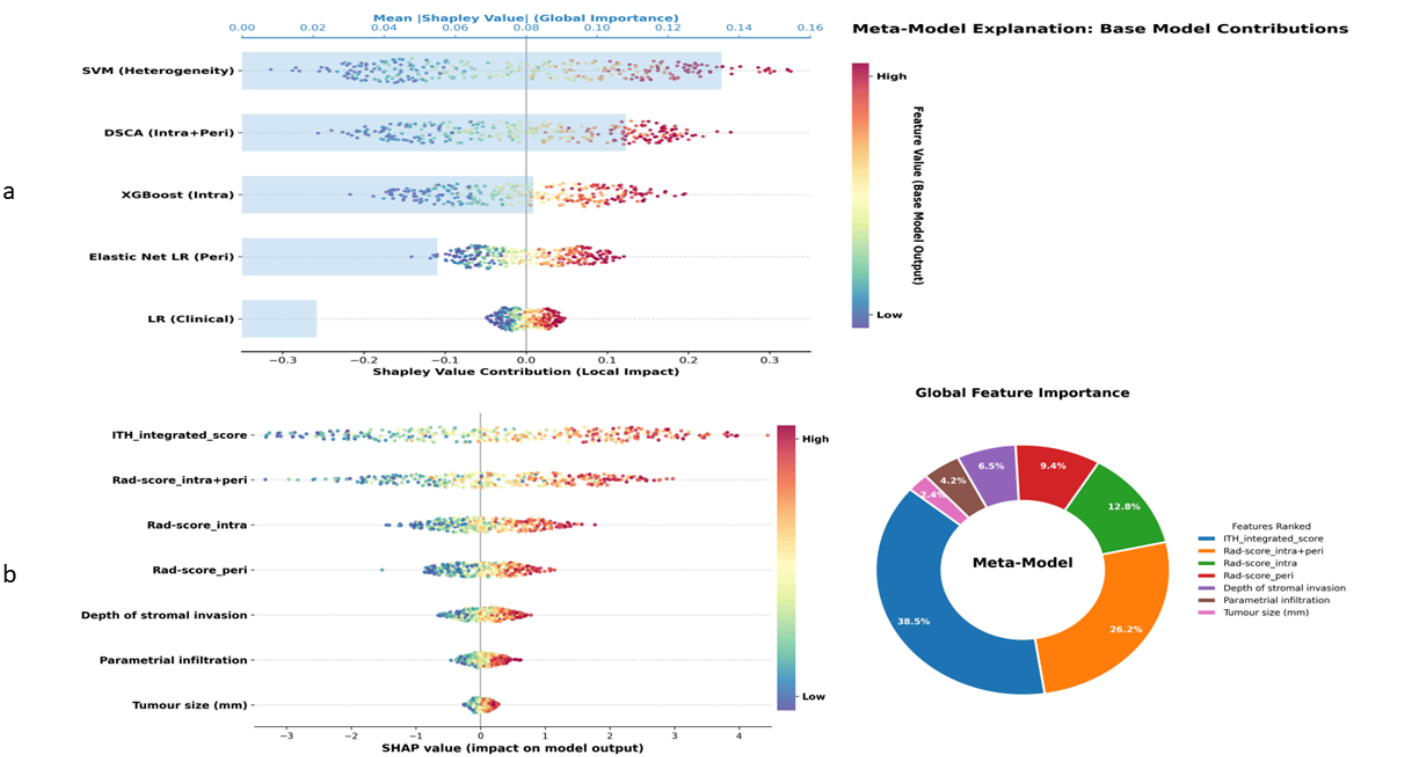


Figure 5

Bee diagram illustrating the contribution of the first layer within the stacking classifier framework.(a) The SVM and DSCA model exhibited the higher mean absolute SHAP value. (b) Global interpretation of the stacking classifier for the second layer, informed by SHAP values. The habitat radiomics score exhibited the highest mean absolute SHAP value.



Figure 6

SHAP analysis with two real-world case example in the force plot. (a) In this force plot, it is evident that the SVM model (0.88), followed by DSCA (0.85), XGBoost (0.8) and Elastic Net LR (0.75), made greater contributions to LNM-positivity prediction. (b) In this force plot, the SVM (0.12) , DSCA (0.15), and XGBoost model (0.2) made negative contributions to the model's prediction, while LR model made positive contributions to LNM-negativity prediction.

Supplementary Files

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

- [SupplementaryInformation.docx](#)