Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Prediction of ERCC1 Expression Level in Nasopharyngeal Carcinoma Based on Multi-sequence MRI Radiomics and Habitat Analysis

Wenjie Zhang
Liuzhou Workers' Hospital, Guangxi Zhuang Autonomous Region
Haixin Huang
Liuzhou Workers' Hospital, Guangxi Zhuang Autonomous Region
Xishan Chen
Liuzhou Workers' Hospital, Guangxi Zhuang Autonomous Region
Ying Lu
1786734840@qq.com

Liuzhou Workers' Hospital, Guangxi Zhuang Autonomous Region

Research Article

Keywords: Habitat analysis, Radiomics, Nasopharyngeal carcinoma, ERCC1, MRI

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11465121/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

Objective

To explore the value of multi-sequence MRI-derived radiomics and habitat analysis for predicting excision repair cross-complementation group 1 (ERCC1) expression in nasopharyngeal carcinoma (NPC).

Methods

In this retrospective study, 171 patients with non-keratinizing undifferentiated NPC without distant metastasis were enrolled. Patients were stratified into high-ERCC1 (n = 75) and low-ERCC1 (n = 96) groups and randomly split into training (n = 136) and test (n = 35) sets at an 8:2 ratio. Radiomic and habitat features were extracted from T1WI, CE-T1WI and T2WI to calculate radscore and habitat radscore. We constructed radiomics, habitat, clinical, radiomics-clinical and habitat-clinical combined models.

Results

A total of 14 radiomic features and 20 habitat features were finally selected. Clinical stage (OR = 1.828, 95%CI: 1.011–3.305, *P* = 0.046) was identified as an independent predictor of high ERCC1 expression. Based on this predictor, five models were constructed: radiomics model, habitat model, radiomics-clinical model, habitat-clinical model and clinical model. Their area under the receiver operating characteristic curve (AUC) values in the training set were 0.825, 0.917, 0.830, 0.922 and 0.605, respectively; the corresponding AUC values in the test set were 0.770, 0.787, 0.770, 0.807 and 0.537. All radiomics-based models yielded higher AUC values than the clinical model (De Long test, all $P < 0.05$). The habitat model outperformed the conventional radiomics model in the training set (De Long test, $P < 0.05$)).

Conclusions

The Multi-sequence MRI-based radiomics and habitat analysis enable effective prediction of ERCC1 expression in NPC patients. The habitat-clinical model achieved the best predictive performance and may support clinical decision-making and prognostic evaluation. The habitat model was superior to conventional radiomics models, suggesting that refined sub-regional segmentation can help boost radiomics prediction accuracy.

Background

According to the GLOBOCAN 2020 database, there were 133 354 newly-diagnosed cases of nasopharyngeal carcinoma (NPC) worldwide. Although the global incidence of NPC is relatively low, China bears nearly half of the global NPC disease burden, which poses a substantial threat to public health[1, 2].At present, the standard therapeutic regimen for NPC consists of platinum-based chemotherapy combined with radiotherapy. Nevertheless, platinum resistance remains a major bottleneck for treatment efficacy and patient prognosis[3].With the advent of a new era for oncological immunotherapy, accumulating clinical studies have reported the safety and efficacy of immunotherapies for NPC[4, 5].Early identification of patients at high-risk of platinum resistance and poor prognosis enables timely immunotherapeutic intervention, which may significantly improve clinical outcomes.

Excision repair cross-complementation group 1 (ERCC1) is a pivotal molecule in the DNA damage-repair pathway. Multiple studies have demonstrated that its expression level is closely correlated with platinum-based chemosensitivity in NPC patients[6–8].Our previous retrospective analysis of 334 NPC patients without distant metastasis revealed that patients with high ERCC1 expression exhibited significantly inferior 4-year overall survival, progression-free survival, and distant metastasis-free survival compared with those with low ERCC1 expression ($P = 0.013$, $P < 0.001$, $P = 0.005$), further confirming the association between elevated ERCC1 expression and adverse prognosis in NPC[9].Accordingly, accurate pre-treatment non-invasive evaluation of ERCC1 expression status is of great clinical importance for stratified management of NPC.

Currently, the gold-standard detection of ERCC1 expression relies on biopsy of tumor lesions to obtain pathological specimens for immunohistochemical staining. However, this invasive approach has inherent limitations. Biopsy only acquires partial tumor tissue and cannot fully reflect the global molecular profile of the entire tumor, thereby introducing sampling bias. Furthermore, biopsy is an invasive procedure carrying potential complications such as hemorrhage and tumor seeding. In addition, ERCC1 expression may dynamically evolve alongside disease progression and therapeutic intervention. Single biopsy fails to capture such temporal molecular alterations, whereas repeated biopsies are impractical in routine clinical practice owing to their invasiveness. Hence, non-invasive tools for comprehensive and dynamic assessment of ERCC1 expression are urgently required.

Radiomics provides a promising strategy for non-invasive prediction of tumor biomarkers. This technique enables high-throughput extraction of quantitative imaging features and characterizes the spatial distribution of intra-tumoral voxels. Numerous studies have confirmed significant associations between radiomic features and molecular biomarkers across various malignancies[10–12].Nevertheless, conventional whole-tumor radiomics extracts features based on the entire tumor volume as a single entity. It ignores spatially distinct biological behaviors within tumors and fails to resolve heterogeneous intra-tumoral subregions, which limits its capacity to dissect intra-tumoral molecular heterogeneity and constrains model performance[13, 14].As an emerging subregional radiomics strategy, habitat analysis compensates for these drawbacks of conventional whole-tumor radiomics. Guided by unsupervised clustering algorithms, habitat analysis automatically partitions tumors at the voxel level and aggregates voxels with similar biological properties into individual habitat units. It captures local

microenvironmental heterogeneity overlooked by whole-tumor analysis and permits refined characterization of intra-tumoral spatial heterogeneity[15].

Conventional radiomics together with habitat analysis has been extensively utilized for non-invasive biomarker prediction in solid tumors including breast cancer and lung cancer, facilitating the progress of precision oncology[16–23].However, studies implementing habitat analysis for predicting platinum-resistance-related biomarkers in NPC remain scarce, and whether habitat features can outperform conventional whole-tumor radiomics for ERCC1 prediction has not yet been clarified. Given the above-mentioned research gaps, this study aimed to construct a predictive model by integrating multi-sequence MRI-derived radiomic and habitat radiomic features together with clinical indicators. We further explored its non-invasive predictive performance for ERCC1 expression levels in patients with nasopharyngeal carcinoma (NPC), to support platinum-resistance risk stratification and individualized therapeutic decision-making for NPC.

Methods

Patient Population

We retrospectively enrolled 256 patients with non-keratinizing undifferentiated NPC without distant metastasis treated at the Fourth Affiliated Hospital of Guangxi Medical University (Liuzhou Workers' Hospital) between June 2020 and June 2023. After applying pre-specified inclusion and exclusion criteria, 171 patients were finally included and stratified into a training set and a test set at an 8:2 ratio.Inclusion criteria: (1) Histopathologically confirmed non-keratinizing undifferentiated nasopharyngeal carcinoma; (2) Absence of distant metastasis and other primary malignancies.Exclusion criteria: (1) Patients who underwent lesion puncture biopsy prior to baseline MRI scanning in our institution; (2) Missing ERCC1 immunohistochemistry results; (3) Poor-quality MRI images prohibiting lesion delineation.(Fig. 1)

The study protocol was approved by the Medical Research Ethics Committee of Liuzhou Workers' Hospital (Approval No. KY2021142). This study was performed in accordance with the Declaration of Helsinki and its subsequent amendments. The requirement for individual informed consent was waived due to the retrospective design.

MRI Equipment and Acquisition Parameters

MRI scans were performed on two scanners. Detailed imaging parameters are summarized in Table S1. Gadopentetate dimeglumine was administered as contrast agent at a dose of 0.1 mmol/kg with an injection rate of 2 mL/s.

ERCC1 Immunohistochemistry Assessment

Two pathologists independently evaluated ERCC1 immunohistochemical slides in a blinded fashion. Concordant results were directly adopted for statistical analysis. Discrepant interpretations were resolved by consensus discussion or adjudication by a third senior pathologist. The proportion of ERCC1-positive tumor cells was scored as 0 (< %), 1 (%-5 %), 2 (6 %50 %), 3 51–75 %), and 4 > 75 %). Staining intensity was scored as 0 (no staining), 1 (weak), 2 (moderate), and 3 (strong). The total score was calculated by summing the proportion score and intensity score. Specimens with total scores of 0–1 were defined as negative (−); 2–3 as weakly positive (+); 4–5 as moderately positive (++); and 6–7 as strongly positive (+++). Cases scored (−)-(+) were categorized as low ERCC1 expression, whereas (++)-(+++) were categorized as high ERCC1 expression[9].

Image Preprocessing and Segmentation

DICOM-format MRI datasets were exported for preprocessing. N4 bias-field correction was performed to mitigate signal inhomogeneity induced by magnetic-field distortion [24],followed by gray-level normalization to reduce inter-scanner and inter-protocol variability for radiomic feature analysis. Using 3D Slicer software (version 5.2.2, https://www.slicer.org), one head-and-neck radiation oncologist with ≥ 3 years of experience manually contoured the primary tumor lesion slice-by-slice on T1WI, CE-T1WI and T2WI sequences to generate whole-tumor volume-of-interest (VOI). Bone, adjacent vessels and retropharyngeal lymph nodes were carefully excluded from VOIs (Figure S1). Thirty randomly selected cases were re-segmented 4 weeks later by the same radiologist and another senior head-and-neck radiation oncologist with ≥ 8 years of experience. Intra-observer and inter-observer agreement were quantified using intraclass correlation coefficients (ICC); ICC > 0.75 was regarded as satisfactory agreement.

Radiomic Feature Extraction

Images were resampled to an isotropic voxel size of 1 mm × 1 mm × 1 mm using linear interpolation for spatial normalization. The Pyradiomics package implemented in Python 3.7 was utilized to extract radiomic features from T1WI, CE-T1WI and T2WI. Apart from original-image features, wavelet-transformed and Laplacian-of-Gaussian-filtered features (σ = 4.0 mm) were computed.

Habitat-Omic Feature Extraction

Gaussian Mixture Model (GMM) clustering was applied to partition intratumoral voxels into biologically homogeneous sub-regions. The Bayesian Information Criterion (BIC) was employed to select the optimal number of clusters *k* within the range from 2 to 8. The *k* value yielding the minimal BIC was chosen as the optimal clustering number. Voxels belonging to identical clusters were assigned unified labels to generate visual habitat maps. Quantitative habitat-omic features were extracted for each segmented tumor sub-region mask (Fig. 2).

Feature Selection

All extracted features were imported into the Yizhun-Darwin online platform (https://darwin-online.yizhun-ai.com). Minimum-maximum normalization was first conducted for feature preprocessing. The op-2 % candidate features were preliminarily retained by percentile-based filtering. The least absolute shrinkage and selection operator (LASSO) logistic regression with five-fold cross-validation was then performed to further reduce dimensionality and identify the most predictive radiomic and habitat-omic features for ERCC1 expression status.

Model Construction

Univariate and multivariate logistic regression analyses were performed to screen independent clinical predictors for high ERCC1 expression to build the clinical model. The selected optimal radiomic and habitat-omic features were used to calculate radscore and habitat radscore, respectively. Logistic regression was adopted to construct the radiomics model and habitat model. Radscore and habitat radscore were separately combined with independent clinical predictors to establish two combined models, whose outputs were visualized as nomograms.

Statistical Analysis

Statistical analyses were performed in R software (version 4.3.1). Receiver-operating-characteristic (ROC) curves, nomograms, calibration curves and decision-curve-analysis (DCA) plots were generated. Student's *t*-test, χ^2-test, Fisher's exact test and logistic regression were applied for clinical-variable comparisons and predictor screening. Model performance was assessed in terms of AUC, accuracy, sensitivity and specificity. The De Long test was used for pairwise AUC comparisons across different models. Calibration curves were plotted to evaluate model goodness-of-fit. Decision-curve analysis was performed to quantify the net clinical benefit of each predictive model.

Results

Baseline Clinical Characteristics

In total, 171 NPC patients were included: 75 patients with high ERCC1 expression and 96 with low ERCC1 expression. The training set contained 136 patients (60 high-expression, 76 low-expression), and the test set contained 35 patients (20 high-expression, 15 low-expression). No statistically significant differences in baseline characteristics existed between training and test cohorts (all $P > 0.05$, Table 1).

Table 1
Baseline clinical characteristics of enrolled NPC patients

Clinical variable	Category	Training set (*n* = 136)	Test set (*n* = 35)	*P* value
ERCC1 expression, *n* (%)	High	76 (55.9)	20 (57.1)	1[a]
	Low	60 (44.1)	15 (42.9)	
Age (years)	Mean ± SD	49.65 ± 12.63	50.00 ± 9.78	0.881[c]
Gender, *n* (%)	Male	99 (72.8)	28 (80.0)	0.514[a]
	Female	37 (27.2)	7 (20.0)	
T-stage, *n* (%)	T1	17 (12.5)	3 (8.6)	0.731[b]
	T2	31 (22.8)	11 (31.4)	
	T3	63 (46.3)	16 (45.7)	
	T4	25 (18.4)	5 (14.3)	
N-stage, *n* (%))	N0	4 (2.9)	1 (2.9)	0.966[b]
	N1	30 (22.1)	9 (25.7)	
	N2	86 (63.2)	21 (60.0)	
	N3	16 (11.8)	4 (11.4)	
Clinical stage, *n* (%)	I	1 (0.7)	0 (0.0)	0.456[b]
	II	11 (8.1)	6 (17.1)	
	III	84 (61.8)	20 (57.1)	
	IVA	40 (29.4)	9 (25.7)	

Note. [a] χ^2-test; [b] Fisher's exact test; [c] Student's *t*-test. NPC, nasopharyngeal carcinoma.

Clinical Predictors and Clinical-Model Development

Univariate logistic regression demonstrated that clinical stage was an independent risk factor for high ERCC1 expression (OR = 1.828, 95%CI: 1.011–3.305, *P* = 0.046). Although T-stage did not reach formal statistical significance (*P* = 0.070), it showed marginal predictive potential. When both T-stage and clinical stage were entered into multivariate logistic regression, neither retained statistical significance (T-stage: OR = 1.219 95 %CI: 0.773–1.952, *P* = 0.397; clinical stage: OR = 1.55, 95 %CI: 0.768–3.180, *P* = 0.222). This phenomenon could be explained by strong collinearity between T-stage and clinical stage[25].Therefore, only clinical stage, which contained more comprehensive prognostic information, was adopted to construct the clinical model (Table S2).

Radiomic-Feature Selection and Radiomics-Model Construction

Initially, 944 radiomic features were extracted from each of the three MRI sequences per patient. After minimum-maximum normalization, percentile-based filtering and LASSO dimensionality reduction, 14 optimal radiomic features were retained. The radscore was calculated from these features for radiomics-model building (Fig. 4a,b).

Habitat-Omic-Feature Selection and Habitat-Model Construction

GMM clustering with BIC criterion yielded the optimal cluster number $k = 2$, dividing tumors into two distinct intratumoral sub-regions (label_1 and label_2). A total of 944 × 3 × 2 habitat-omic features were extracted for each patient. Following the same preprocessing and LASSO-based feature-selection pipeline, 20 high-value habitat-omic features were screened. The habitat radscore was computed to develop the habitat model (Fig. 3c,d).

Combined Prediction Models

Radscore and habitat radscore were separately integrated with clinical stage to generate the radiomics-clinical combined model and habitat-clinical combined model, which were visualized as nomograms (Fig. 4). Higher total points in nomograms corresponded to higher predicted probability of high ERCC1 expression.

Predictive Performance of All Models

ROC-curve analyses revealed that in the training set, AUC values for clinical model, radiomics model, habitat model, radiomics-clinical model and habitat-clinical model were 0.605, 0.825, 0.917, 0.830 and 0.922, respectively. Corresponding AUC values in the test set were 0.537, 0.770, 0.787, 0.770 and 0.807 (Table 2, Fig. 5a,b). All radiomics-related models achieved significantly higher AUC than the clinical model (De Long test, all $P < 0.05$). In the training cohort, the habitat-model AUC significantly exceeded that of the conventional radiomics model (De Long test, $P < 0.05$). Although combined models achieved slightly elevated AUC relative to single-modality radiomics-based models, the differences lacked statistical significance (De Long test, $P > 0.05$, Table S3). Calibration curves demonstrated favorable goodness-of-fit for radiomics-based and habitat-based models (Fig. 5e,f). Decision-curve analysis indicated that radiomics-derived models provided superior net clinical benefit compared with the standalone clinical model across reasonable threshold probability ranges (Fig. 5c,d).

Table 2
Predictive performance of clinical, radiomic, habitat and combined models

Training set	Model	AUC (95%CI)	Accuracy (%)	Sensitivity (%)	Specificity (%)
	Clinical model	**0.605(0.519–0.692)**	**0.441**	**0.9**	**0.079**
	Radiomics model	0.825(0.755–0.895)	0.765	0.733	0.789
	Habitat model	0.917(0.872–0.962)	0.873	0.8	0.908
	Radiomics-clinical modelical model	0.83(0.761–0.899)	0.772	0.6	0.908
	Habitat-clinical model	0.922(0.879–0.965)	0.831	0.817	0.868
Test set	Clinical model	0.537(0.353–0.721)	0.429	0.8	0.15
	Radiomics model	0.77(0.609–0.931)	0.714	0.8	0.65
	Habitat model	0.787(0.63–0.943)	0.714	0.667	0.8
	Radiomics-clinical model	0.77(0.608–0.932)	0.743	0.8	0.7
	Habitat-clinical model	0.807(0.655–0.959)	0.684	0.867	0.7

Discussion

This study constructed conventional radiomic models, habitat models, and corresponding clinical combined models based on multi-sequence MRI data to systematically evaluate their value for noninvasively predicting ERCC1 expression in nasopharyngeal carcinoma (NPC) patients. The results demonstrated that all omics models, except the purely clinical model, could effectively distinguish patients with high and low ERCC1 expression, providing objective imaging evidence for individualized treatment and prognostic evaluation of NPC. Notably, the habitat model exhibited superior predictive performance compared with conventional whole-tumor radiomics, confirming that refined subregional heterogeneity analysis can substantially improve the predictive accuracy of tumor molecular phenotyping.

A total of 14 optimal radiomic features and 20 habitat features were finally screened in this study, among which texture features were predominant. These feature distributions were consistent with those reported in previous NPC biomarker studies[26],ndicating that texture parameters serve as core imaging

indicators for quantifying tumor microstructures and intratumoral heterogeneity, and can indirectly reflect differential ERCC1 expression. Clinical correlation analysis revealed that clinical stage was an independent risk factor for high ERCC1 expression (OR = 1.828, 95%CI: 1.011–3.305, P = 0.046). Patients with advanced clinical stages were more likely to present high ERCC1 expression, which was in line with previous findings[9, 27].Nevertheless, the standalone clinical staging model showed limited predictive efficacy (training set AUC = 0.605; test set AUC = 0.537). The fundamental reason is that clinical staging only reflects macroscopic anatomical tumor invasion and fails to capture microscopic intratumoral heterogeneity and molecular phenotypic characteristics, which restricts its capacity for precise risk stratification and highlights the indispensable value of auxiliary radiomic models.

Most previous radiomics studies for tumor-biomarker prediction have mainly focused on globally prevalent malignancies[28–30].By contrast, investigations on relatively rare cancers such as nasopharyngeal carcinoma remain scarce, and existing studies are largely confined to Ki-67 prediction. In the present study, we innovatively selected ERCC1 expression, a biomarker of great importance for clinical decision-making, as the outcome variable. Models constructed using radiomic features from T1WI, T2WI and CE-T1WI achieved favorable performance, with an AUC of 0.825 in the training set and 0.77 in the test set, which was in line with the findings reported by Bi et al[31].Bi et al. retrospectively analyzed MRI data from 128 patients with sinonasal malignancies and built a model based on T1WI, T2WI and CE-T1WI sequences for Ki-67 prediction. Their model achieved an AUC of 0.852 (95%CI: 0.733–0.971) and an accuracy of 86.3% in the test cohort; however, their model only incorporated radiomic features without integrating clinical variables. Wang et al. further combined multi-sequence MRI radiomic signatures with two independent clinical predictors, namely plasma EBV-DNA load and tumor enhancement degree[26].Their combined model showed superior performance for predicting high Ki-67 expression in NPC (training-set AUC = 0.887, test-set AUC = 0.86) compared with the standalone clinical model, yet it did not significantly improve the performance of the pure radiomics model (De Long test, $P > 0.05$). The optimal features screened in our study were derived from all three MRI sequences (T1WI, T2WI, and CE-T1WI), demonstrating that multi-sequence fusion enables multi-dimensional mining of tumor imaging information. Conventional radiomic models significantly outperformed pure clinical models (all $P < 0.05$), further verifying the unique advantages of radiomic features in capturing microscopic tumor heterogeneity. The radiomics-clinical combined model showed a marginal improvement in predictive efficacy compared with the standalone radiomic model, suggesting complementary potential between imaging phenotypes and clinical variables, which deserves further validation with larger sample sizes.

The most innovative and core finding of this study is that habitat analysis can break through the inherent limitations of conventional whole-tumor radiomics and significantly improve the predictive accuracy of ERCC1 expression in NPC. Traditional whole-tumor radiomics averages imaging features across the entire tumor volume, ignoring biological differences among intratumoral subregions and masking subtle microenvironmental heterogeneity, thereby leading to information loss and limited predictive performance. In contrast, habitat analysis adopts voxel-level unsupervised clustering to automatically partition tumors into multiple biologically homogeneous subregions. This strategy enables refined

characterization of spatial intratumoral heterogeneity and captures subtle regional imaging differences that cannot be identified by conventional radiomics, thereby more accurately reflecting the spatial distribution of tumor microenvironment and molecular phenotypes. In the present study, the habitat model achieved excellent predictive performance with an AUC of 0.917 in the training set and 0.787 in the test set, strongly demonstrating the superiority of refined subregional modeling for NPC molecular biomarker prediction.

The robust predictive capacity of habitat analysis has been validated in multiple solid tumors. Chen et al. successfully distinguished breast cancer HER2 expression status through habitat heterogeneity quantification and achieved stable performance across multiple cohorts[32].Wu et al. constructed CT-based habitat models for EGFR mutation screening in stage □ non-small cell lung cancer and obtained better predictive efficacy than conventional models[22].providing solid theoretical and experimental support for our results. In this study, we adopted the K-means clustering algorithm, the most widely recognized and stable unsupervised segmentation method in current habitat radiomics research[33, 34].This algorithm achieves objective and stable subregional segmentation based on multi-sequence MRI voxel similarity, independent of spatial voxel location, and effectively separates functionally heterogeneous tumor compartments. Based on existing imaging-pathological evidence, we speculated that the two segmented habitat subregions correspond to highly perfused, metabolically active tumor areas and hypoperfused, necrosis-dominated hypoxic areas, respectively. Differences in cellular density, microvascular perfusion, and hypoxia among different subregions directly regulate the activity of DNA damage repair pathways, ultimately resulting in spatial heterogeneity of ERCC1 expression[35].Although precise spatial registration between imaging subregions and pathological slices was not performed in this study, the capacity of habitat features to reflect tumor genotypes, microenvironmental alterations, and biological phenotypes has been widely confirmed, which fundamentally explains the higher accuracy and robustness of habitat models compared with conventional whole-tumor radiomics.

This study has several limitations. First, certain critical clinical indicators (e.g., EBV-DNA load) were excluded due to incomplete data, and future studies incorporating more clinical parameters may further optimize model performance. Second, this was a single-center retrospective study with a relatively limited sample size; multi-center external validation is required to verify the generalizability of the model. Third, manual lesion segmentation is challenging due to the complex anatomical structure of the nasopharynx, and subjective bias may affect model repeatability, which can be improved by automated segmentation algorithms in future research. Fourth, only logistic regression was adopted for model construction; future studies may introduce diverse machine learning algorithms to further explore the potential of habitat features.

Conclusion

In conclusion, the multi-sequence MRI-based conventional radiomic model, habitat model, and their corresponding clinical combined models can effectively and noninvasively predict ERCC1 expression in NPC patients. Among them, the habitat-clinical combined model yields the optimal predictive efficacy,

providing valuable evidence for platinum resistance risk stratification and individualized NPC treatment. Most importantly, this study demonstrates that habitat subregional analysis enables finer quantification of tumor heterogeneity and outperforms conventional whole-tumor radiomics, offering a novel and promising imaging strategy for future precision biomarker research in NPC.

Declarations

Ethical approval

The study protocol was approved by the Medical Research Ethics Committee of Liuzhou Workers' Hospital (Approval No. KY2021142). This study was performed in accordance with the Declaration of Helsinki and its subsequent amendments. The requirement for individual informed consent was waived due to the retrospective design.

Consent to participate

All patient data were anonymized. The requirement for informed consent was waived because of the retrospective nature of this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

The authors declare that this study and/or its publication received financial support. This work was supported by the Guangxi Key Research and Development Program (Grant No. Guike AB22035026) and the Natural Science Foundation of Guangxi Zhuang Autonomous Region (Grant No. 2024GXNSFAA010116).

Author Contribution

ZWJ: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. LY: Formal analysis, Writing – review & editing. HXX: Formal analysis, Writing – review & editing. CXS: Formal analysis, Writing – review & editing.

Data availability

No datasets were generated or analysed during the current study.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Cancer J Clin. 2021;71(3):209–49.
2. Chen YP, Chan ATC, Le QT, et al. Nasopharyngeal carcinoma. Lancet (London England). 2019;394(10192):64–80.
3. Chen YP, Ismaila N, Chua MLK, et al. Chemotherapy in Combination With Radiotherapy for Definitive-Intent Treatment of Stage II-IVA Nasopharyngeal Carcinoma: CSCO and ASCO Guideline. J Clin oncology: official J Am Soc Clin Oncol. 2021;39(7):840–59.
4. Xu JY, Wei XL, Wang YQ, et al. Current status and advances of immunotherapy in nasopharyngeal carcinoma. Therapeutic Adv Med Oncol. 2022;14:17588359221096214.
5. Huang H, Yao Y, Deng X et al. Immunotherapy for nasopharyngeal carcinoma: Current status and prospects (Review). Int J Oncol 2023, 63(2).
6. Bišof V, Zajc Petranović M, Rakušić Z et al. The prognostic and predictive value of excision repair cross-complementation group 1 (ERCC1) protein in 1288 patients with head and neck squamous cell carcinoma treated with platinum-based therapy: a meta-analysis. *European archives of oto-rhino-laryngology: official journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS) : affiliated with the German Society for Oto-Rhino-Laryngology - Head and Neck Surgery* 2016, 273(9):2305–2317.
7. Xu S, Yu Y, Rong J, et al. Expression of BRCA1 and ERCC1 as predictive clinical outcome after radiochemotherapy in patients with locoregionally moderate-advanced nasopharyngeal carcinoma. Oncotarget. 2017;8(19):31355–67.
8. Sabatella M, Thijssen KL, Davó-Martínez C, et al. Tissue-Specific DNA Repair Activity of ERCC-1/XPF-1. Cell Rep. 2021;34(2):108608.
9. Lu Y, Huang H, Kang M, et al. Combined Ki67 and ERCC1 for prognosis in non-keratinizing nasopharyngeal carcinoma underwent chemoradiotherapy. Oncotarget. 2017;8(51):88552–62.
10. Gillies RJ, Kinahan PE, Hricak H, et al. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563–77.
11. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J nuclear medicine: official publication Soc Nuclear Med. 2020;61(4):488–95.
12. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328–38.
13. Lord AC, D'Souza N, Shaw A, et al. MRI-Diagnosed Tumor Deposits and EMVI Status Have Superior Prognostic Accuracy to Current Clinical TNM Staging in Rectal Cancer. Ann Surg. 2022;276(2):334–44.

14. Yang YS, Feng F, Qiu YJ, et al. High-resolution MRI-based radiomics analysis to predict lymph node metastasis and tumor deposits respectively in rectal cancer. Abdom Radiol (New York). 2021;46(3):873–84.
15. Jia P, Li Y, Li H, et al. Habitat radiomics assists radiologists in accurately diagnosing lymph node metastasis of adenocarcinoma of the esophagogastric junction. Insights into imaging. 2025;16(1):90.
16. Ramtohul T, Djerroudi L, Lissavalid E, et al. Multiparametric MRI and Radiomics for the Prediction of HER2-Zero, -Low, and -Positive Breast Cancers. Radiology. 2023;308(2):e222646.
17. Mingjie CY. Preoperative prediction of Ki-67 expression status in breast cancer using a combined intra-tumoral and peritumoral radiomics model based on DCE-MRI. Magn Reson Imaging. 2022;13(10):132–7.
18. Sang L, Liu Z, Huang C, et al. Multiparametric MRI-based radiomics nomogram for predicting the hormone receptor status of HER2-positive breast cancer. Clin Radiol. 2024;79(1):60–6.
19. Wang Y, Wan Q, Xia X, et al. Value of radiomics model based on multi-parametric magnetic resonance imaging in predicting epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. J Thorac disease. 2021;13(6):3497–508.
20. Shi W, Yang Z, Zhu M, et al. Correlation between PD-L1 expression and radiomic features in early-stage lung adenocarcinomas manifesting as ground-glass nodules. Front Oncol. 2022;12:986579.
21. Wang Y, Zhang Y, Liu Z, et al. Multiparametric MRI-based habitat analysis integrating deep learning and radiomics for predicting preoperative Ki-67 expression level in breast cancer. BMC Med Imaging. 2026;26(1):80.
22. Wu J, Meng H, Zhou L, et al. Habitat radiomics and deep learning fusion nomogram to predict EGFR mutation status in stage I non-small cell lung cancer: a multicenter study. Sci Rep. 2024;14(1):15877.
23. Ding M, He T, Yu J, et al. Prediction of intracranial response to PD-1/PD-L1 inhibitors therapy in brain metastases originating from non-small cell lung cancer using habitat imaging and peritumoral radiomics: a multicenter study. Front Oncol. 2025;15:1657290.
24. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310–20.
25. Varady NH, Pareek A, Eckhardt CM, et al. Multivariable regression: understanding one of medicine's most fundamental statistical tools. Knee Surg sports Traumatol arthroscopy: official J ESSKA. 2023;31(1):7–11.
26. Wang Zhuo L, Shili Z, Shaoru, et al. Predicting Nasopharyngeal Carcinoma Tumor Cell Proliferative Activity Using Multi-Parametric MRI Radiomics Combined with Clinical Imaging Features. Magn Reson Imaging. 2022;13(11):30–6.
27. Li W, Qi S. Association between expression of nucleotide excision repair cross-complementation group 1 and efficacy of platinum-based drugs in locally advanced nasopharyngeal carcinoma. J Clin Otorhinolaryngol Head Neck Surg. 2015;29(02):144–6.

28. Hong D, Xu K, Zhang L, et al. Radiomics Signature as a Predictive Factor for EGFR Mutations in Advanced Lung Adenocarcinoma. Front Oncol. 2020;10:28.
29. Hao P, Deng BY, Huang CT, et al. Predicting anaplastic lymphoma kinase rearrangement status in patients with non-small cell lung cancer using a machine learning algorithm that combines clinical features and CT images. Front Oncol. 2022;12:994285.
30. Liu Tingting L, Jialu L, Jianjuan et al. The clinical application value of multi-parameter MRI imageomics in evaluating the HER-2 expression status of invasive breast cancer. J Nanjing Med Univ (Natural Sci Edition) 2024(02):218–27.
31. Bi S, Li J, Wang T, et al. Multi-parametric MRI-based radiomics signature for preoperative prediction of Ki-67 proliferation status in sinonasal malignancies: a two-centre study. Eur Radiol. 2022;32(10):6933–42.
32. Chen H, Liu Y, Zhao J, et al. Quantification of intratumoral heterogeneity using habitat-based MRI radiomics to identify HER2-positive, -low and -zero breast cancers: a multicenter study. Breast cancer research: BCR. 2024;26(1):160.
33. Yang Y, Han Y, Zhao S, et al. Spatial heterogeneity of edema region uncovers survival-relevant habitat of Glioblastoma. Eur J Radiol. 2022;154:110423.
34. Du T, Zhao H. Habitat Analysis of Breast Cancer-Enhanced MRI Reflects BRCA1 Mutation Determined by Immunohistochemistry. *BioMed research international* 2022, 2022:9623173.
35. Liu J, Xu M, Ren J, et al. Synthetic MRI, multiplexed sensitivity encoding, and BI-RADS for benign and malignant breast cancer discrimination. Front Oncol. 2022;12:1080580.

Figures

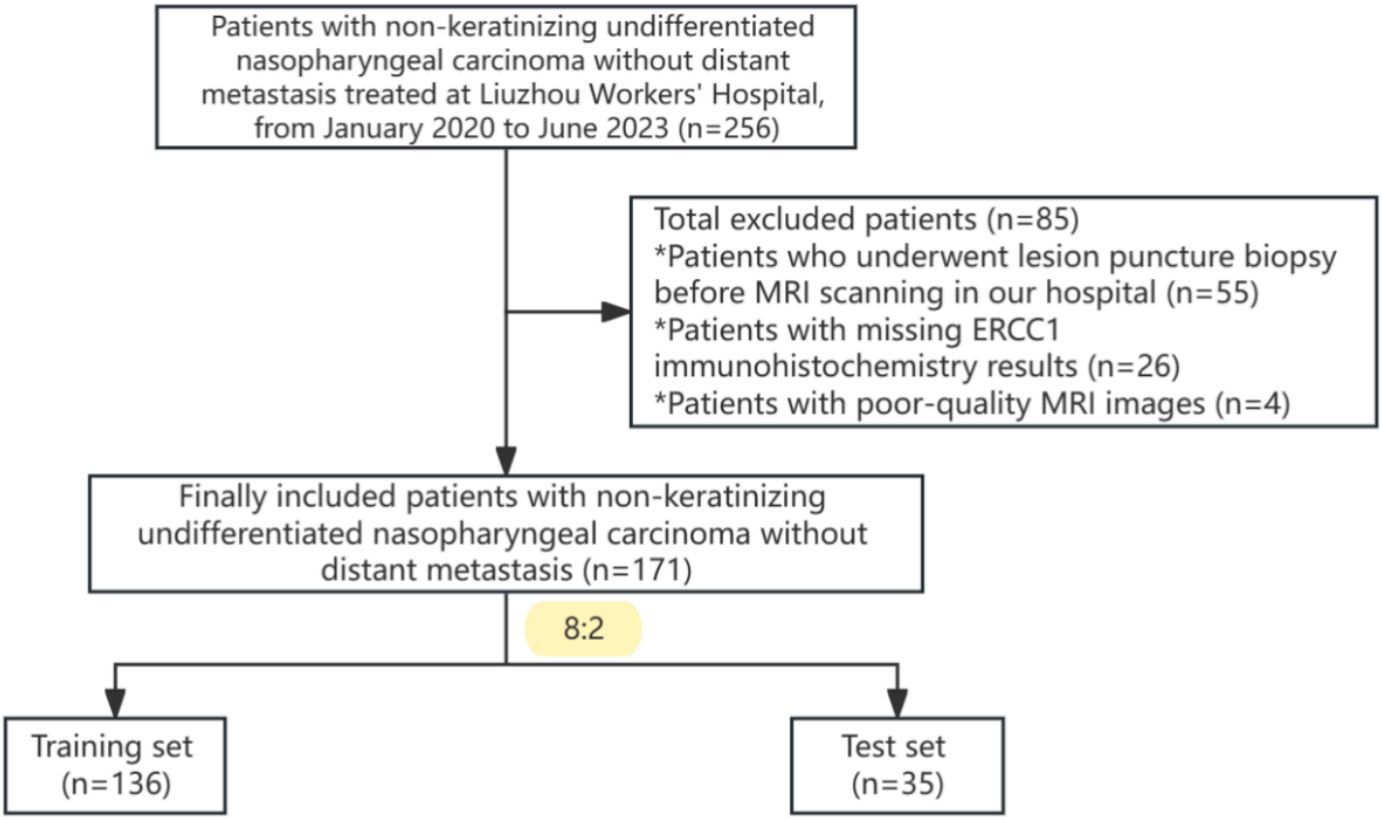


Figure 1

Flowchart of patient inclusion and exclusion criteria.

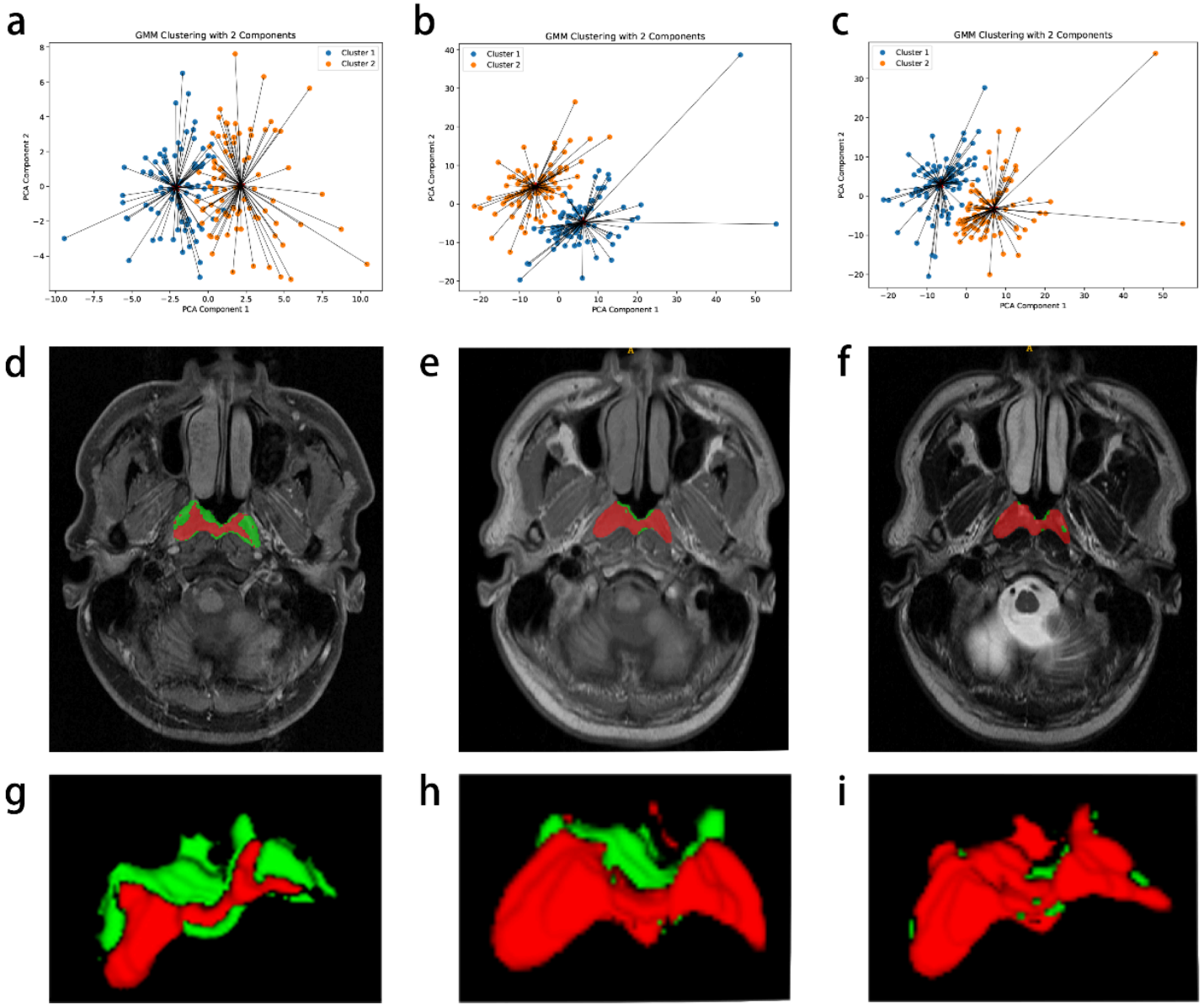


Figure 2

(a), (d), and (g) show the optimal clustering and maps of tumor subregions on CE-T1WI; (b), (e), and (h) show the optimal clustering and maps of tumor subregions on T1WI; (c), (f), and (i) show the optimal clustering and maps of tumor subregions on T2WI.

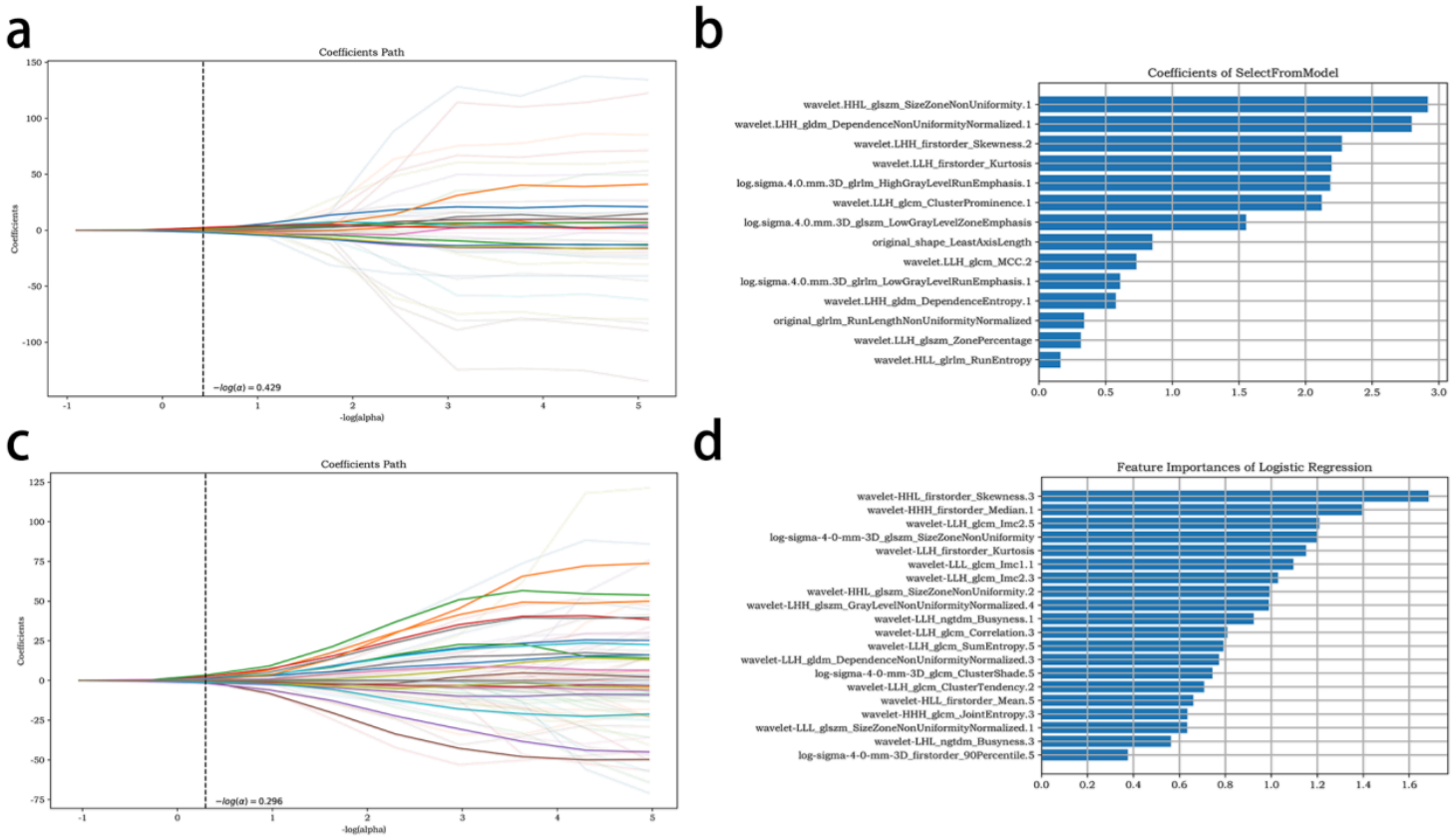


Figure 3

(a) Selection of the optimal parameter -log(α)=0.429 for the radiomics model using LASSO regression with 5-fold cross-validation; (b) The 14 optimal radiomic features for predicting ERCC1 expression status in patients with nasopharyngeal carcinoma (NPC); (c) Selection of the optimal parameter -log(α)=0.296 for the habitat model using LASSO regression with 5-fold cross-validation; (d) The 20 optimal habitat features for predicting ERCC1 expression status in patients with NPC.

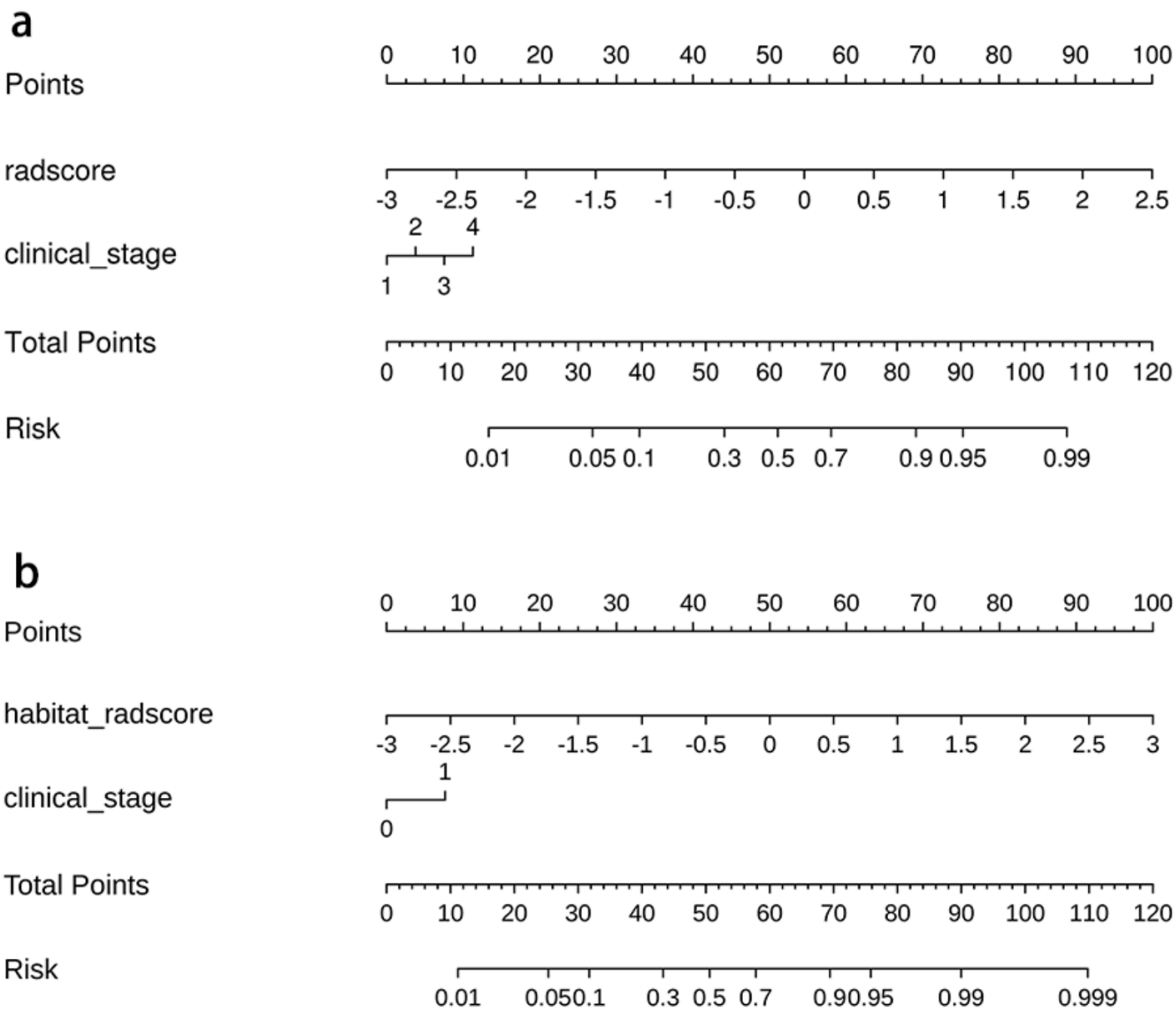


Figure 4

(a) Nomogram of the radiomics-clinical combined model; (b) Nomogram of the habitat-clinical combined model for predicting high ERCC1 expression in patients with nasopharyngeal carcinoma.

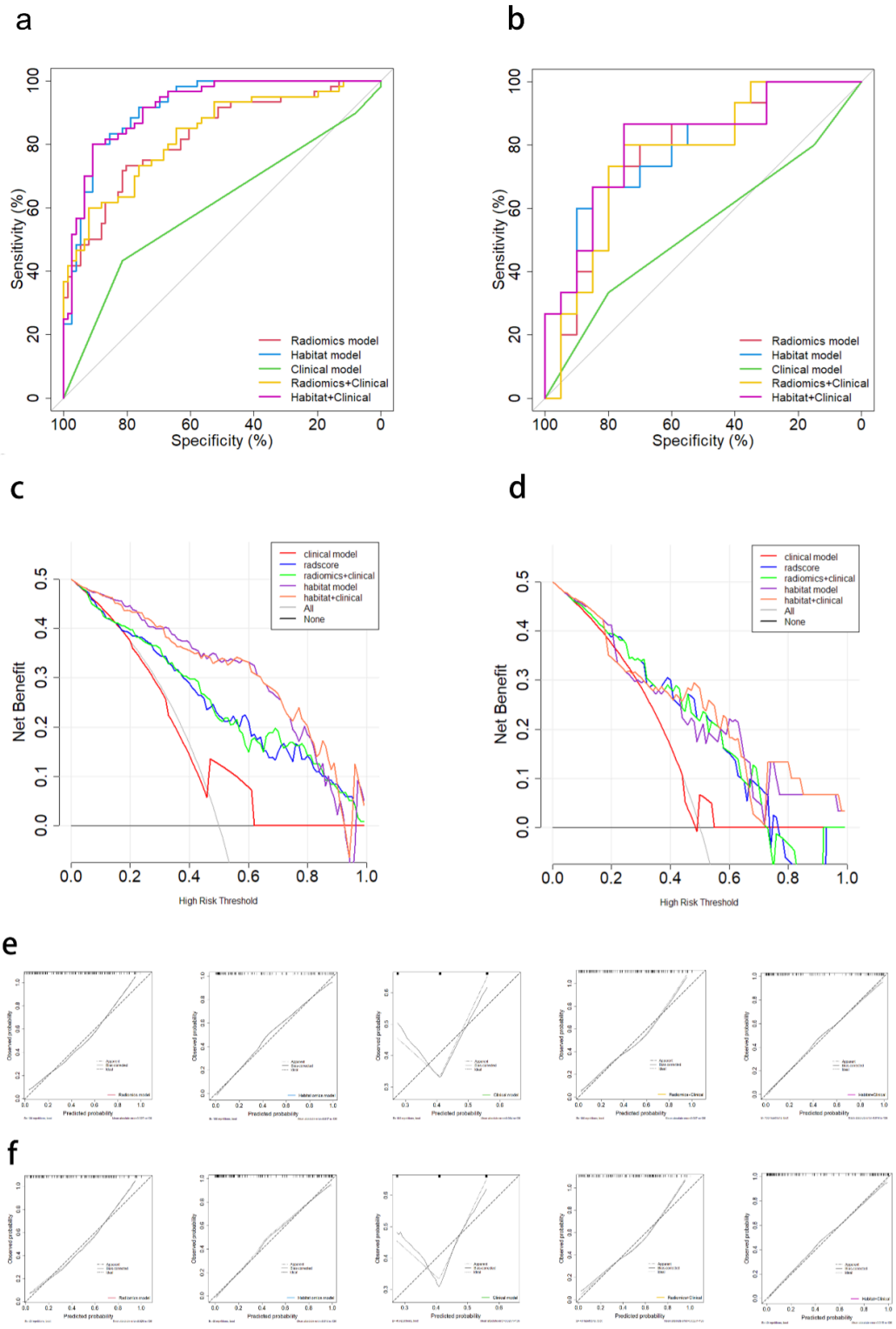


Figure 5

(a) ROC curves of all models in the training set; (b) ROC curves of all models in the test set; (c) Decision-curve analysis (DCA) curves of all models in the training set; (d) Decision-curve analysis (DCA) curves of all models in the test set; (e) Calibration curves of all models in the training set; (f) Calibration curves of all models in the test set.

Supplementary Files

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

- supplementarymaterials.docx