Clinical & Experimental Metastasis (2025) 42:30
https://doi.org/10.1007/s10585-025-10349-y

RESEARCH PAPER

Deep learning MRI-based radiomic models for predicting recurrence in locally advanced nasopharyngeal carcinoma after neoadjuvant chemoradiotherapy: a multi-center study

Chunmiao Hu[1] · Congrui Xu[3] · Jiaxin Chen[3] · Yiling Huang[3] · Qingcheng Meng[4] · Zhian Lin[5] · Xinming Huang[6] · Li Chen[2]

Received: 25 July 2024 / Accepted: 28 April 2025

Abstract

Local recurrence and distant metastasis were a common manifestation of locoregionally advanced nasopharyngeal carcinoma (LA-NPC) after neoadjuvant chemoradiotherapy (NACT). To validate the clinical value of MRI radiomic models based on deep learning for predicting the recurrence of LA-NPC patients. A total of 328 NPC patients from four hospitals were retrospectively included and divided into the training($n=229$) and validation ($n=99$) cohorts randomly. Extracting 975 traditional radiomic features and 1000 deep radiomic features from contrast enhanced T1-weighted (T1WI+C) and T2-weighted (T2WI) sequences, respectively. Least absolute shrinkage and selection operator (LASSO) was applied for feature selection. Five machine learning classifiers were conducted to develop three models for LA-NPC prediction in training cohort, namely Model I: traditional radiomic features, Model II: combined the deep radiomic features with Model I, and Model III: combined Model II with clinical features. The predictive performance of these models were evaluated by receive operating characteristic (ROC) curve analysis, area under the curve (AUC), accuracy, sensitivity and specificity in both cohorts. The clinical characteristics in two cohorts showed no significant differences. Choosing 15 radiomic features and 6 deep radiomic features from T1WI+C. Choosing 9 radiomic features and 6 deep radiomic features from T2WI. In T2WI, the Model II based on Random forest (RF) (AUC=0.87) performed best compared with other models in validation cohort. Traditional radiomic model combined with deep radiomic features shows excellent predictive performance. It could be used assist clinical doctors to predict curative effect for LA-NPC patients after NACT.

Keywords Deep learning · Radiomic features · Locoregionally advanced nasopharyngeal carcinoma · Recurrence prediction · Neoadjuvant chemotherapy · MRI · Multi-center study

✉ Li Chen
17138935@qq.com

[1] Department of Radiology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, Fuzhou Fujian 350014, China

[2] Department of Mathematics and Computer, School of Arts and Sciences, Fujian Medical University, University Town, No 1 North Xuefu Road, Fuzhou Fujian, 350122, China

[3] Department of Clinical Medicine, Fujian Medical University, Fuzhou 350122, Fujian, China

[4] Department of Radiology, The Affiliated Tumor Hospital of Zhengzhou University & Henan Tumor Hospital, Zhengzhou Henan 450000, China

[5] Department of Radiation Oncology, Zhongshang Hospital Xiamen University, Xiamen Fujian 361000, China

[6] Department of Radiology, Fujian Medical University Union Hospital, Fuzhou Fujian 350001, China

Abbreviations

NPC	Nasopharyngeal carcinoma
LA-NPC	Locoregionally advanced nasopharyngeal carcinoma
NACT	Neoadjuvant chemoradiotherapy
CCRT	Concurrent radiotherapy
IC	Induction chemotherapy
IMRT	Intensity modulated radiation therapy
MRI	Magnetic resonance imaging
AUC	Area under curve
T1WI+C	Contrast enhanced T1-weighted image
T2WI	T2-weighted MRI imaging
ROI	Region of interest
LASSO	Least absolute shrinkage and selection operator

Published online: 15 May 2025

ROC	Receiver operator characteristic
ResNet	Residual neural networks
SVM	Support vector machine
RF	Random forest
AdaBoost	Adaptive boosting
KNN	K-nearest neighbor
NB	Naive Bayes

Introduction

Nasopharyngeal carcinoma (NPC) is usually malignant tumor of the head and neck prevails in Southeast Asia area. There were 120,416 new cases of NPC were recorded, resulting in 73,476 deaths by world area in 2022 [1]. Radiotherapy is the primary treatment for nasopharyngeal carcinoma, and for patients with locoregionally advanced nasopharyngeal carcinoma (LA-NPC), neoadjuvant chemotherapy (NACT) combined with concurrent radiotherapy (CCRT) and induction chemotherapy (IC) can improve progression-free survival [2–4]. In the era of intensity modulated radiation therapy (IMRT), local recurrence and distant metastasis of NPC patients became the main reasons for treatment failure [5–7]. After receiving radical treatment, about 20% of NPC patients have local recurrence and distant metastasis [8]. So it is critical to assess LA-NPC patients who are at higher risk of progression after they received the NACT.

Magnetic resonance imaging (MRI) is a commonly used imaging method to stage and evaluate the curative effect of nasopharyngeal carcinoma. MRI has high soft tissue resolution, can clearly show the anatomic details, the scope of lesion infiltration, and has high diagnostic efficiency for cervical lymph node metastasis and local residual or recurrence [9].

Radiomics provides a quantified information of the tumors in non-invasive way. It can realize the visualization of tumor heterogeneity and help improve the ability of medical imaging in the diagnosis, prognosis prediction and efficacy monitoring. Several studies have reported that the radiomics based on MRI have the potential clinical value, the traditional radiomic features of the tumors could be used to predicting the preoperative assessment, treatment response and survival and prognosis [10–13]. Although these reports figured that traditional radiomic features can adequately express tumor heterogeneity, the average predictive ability needs to be improved. As everyone knows, deep learning methods performed well in medical image analysis. Recent studies focused on the predictive ability of deep learning methods in other cancers [14, 15]. Few study have analyzed the deep radiomics features of LA-NPC patients' tumors, and combined these features with traditional radiomic features. The purpose of this study is to establish MRI radiomic models that contained traditional and deep radiomic features, explore and evaluate the clinical value of these radiomic models for predicting the recurrence of LA-NPC patients after NACT treatment.

Materials and methods

Data

The Ethics Committee of Fujian Cancer Hospital approved this retrospective study and the requirement for informed patient consent was waived. A total of 340 confirmed LA-NPC patients received NAC from January 2014 to July 2017 were enrolled (Hospital 1, 268 cases, outside hospitals: Hospital 2, 29 cases, Hospital 3, 33 cases and Hospital 4, 10 cases). Before received the same treatment and after the end of the second cycle of NACT, all patients underwent MRI examinations. Therapeutic regimen were detailed in Supplementary file(Sect. 1). Contrast enhanced T1-weighted image (T1WI+C) and T2-weighted image (T2WI) sequences were collected after the end of NACT two cycles. All patients accepted systemic examination after the 5-year completion of treatment.

In this study, a total of 328 LA-NPC patients were included based on the inclusion and exclusion criteria. Patients inclusion criteria: (1) Locally advanced nasopharyngeal carcinoma (stage III-IVa) confirmed by pathology. (2) Absent treatment before examination. (3) Re-examined with MRI after the end of two cycles of neoadjuvant chemotherapy. (4) Nasopharyngeal MRI examination before and after complete treatment. (5) Patients without recurrence have whole images over a 5-year follow-up period. (6) Local recurrence is confirmed by pathology, and distant metastasis is confirmed by pathology or multiple imaging examinations. These recurrence were contained the local recurrence and distant metastasis. So these patients with local recurrence or distant metastasis within 5 years were divided into the recurrence cohort, and those without local recurrence or distant metastasis were divided into the non-recurrence cohort. Patient exclusion criteria: (1) The image sequence is incomplete and the image quality does not meet the diagnostic criteria. (2) The patient lost to follow-up.

This study only collected age, gender, T-stage, N-stage features, because of the missing data in some hospital. Although radiotherapy dose is an important clinical parameter, all patients in our study received a similar dosing regimen, as described in the supplementary file (Sect. 1), so the dose parameters did not show significant variability in our cohort and were not included in our clinical features. Finally, a total of 222 cases in the 5-year non-recurrence cohort and 106 cases in the 5-year recurrence cohort were

included. The patients were randomly divided into a training cohort ($n=229$) and a validation cohort ($n=99$) at 7:3 ratio.

The following workflow of radiomics analysis process was consisted of four parts as shown in Fig. 1: (I) Image acquisition and segmentation; (II) Feature extraction; (III) Feature selection; (IV) Radiomic model construction and validation.

Image acquisition and segmentation

MRI examinations were performed on 3.0-T scanners (Achieva 3.0T TX, Philips, Germany or GE Discovery MR 750, General Electric Medical Systems, USA). Axial contrast enhanced T1-weighted (T1WI+C) and T2-weighted images (T2WI) were gained from the examinations. Digital imaging and communications in medicine (DICOM) images were collected from picture archiving and communication system (PACS). The protocols of T1WI+C and T2WI sequences with parameters used is detailed in Supplementary file(Sect. 2).

The regions of interest (ROIs) were manually outlined the contour of the entire tumors slice by slice on cross section of MRI images, including necrosis, vessels, etc. All three-dimensional segmentation were blindly executed by two radiologists (radiologist A with 10 years of experience, radiologist B with 20 years of experience) through ITK-SNAP software (v.3.4.0, http://www.itksnap.org), an open-source software for medical image segmentation [16]. Thirty cases were randomly selected, and the ROI was delineated once by radiologist A, and again delineated one week later to evaluate the intraclass coefficient correlation (ICC). The radiologist B only delineated the ROI once to evaluate the ICC between the two groups. When ICC>0.75, it would be considered to good consistency.

Feature extraction and selection

Feature extraction was implemented by the PyRadiomics package in Python software (3.6.4) [17]. To account for potential variations in imaging protocols across different centers, we performed intensity normalization across all original DICOM imaging data. After standardizing images, a total of 975 traditional radiomic features were extracted form T1WI+C and T2WI sequences for each patients, respectively. These radiomic features contained original image features, first-order features, grayscale-dependent matrix features (GLDM), grayscale run matrix feature (GLRLM), grayscale size region matrix features (GLSZM) and wavelet features. The type and quantity of features were detailed in Supplementary file(Sect. 3).

Previous studies indicated that residual neural networks (ResNet) can be used to diagnose the lesions or extract the radiomic features [18–19]. In this study, a pre-trained ResNet50 model through transfer learning was utilized to extract the deep radiomic features on the T1WI+C and

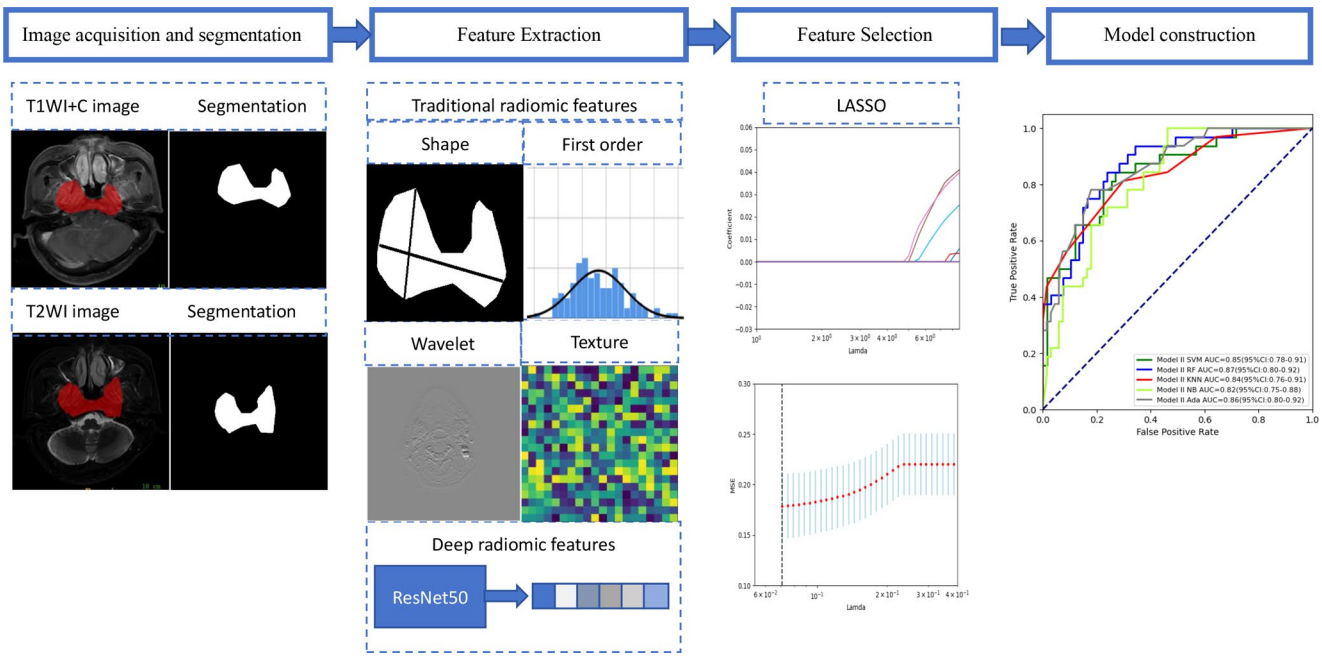

Fig. 1 The workflow of the multiple radiomic models construction and validation. Image acquisition and segmentation: collected two sequences images and segment the tumor by two radiologists. Feature extraction: extracted traditional and deep radiomic features from these images. Feature selection: used the LASSO method to select the optimal features. Model construction: generated combined radiomic model based on selected radiomic features and validate the predictive performance

T2WI sequence images. This pre-trained ResNet50 model contained 50 convolutional layers plus fully connected layers based on ResNet architecture. The ResNet50 model was used to extract 1000 deep radiomic features from T1WI+C and T2WI sequences, respectively. And named with F+numbers, because of their unexplainability.

Additionally, to further mitigate center-specific biases in radiomic features, ComBat correction was applied, a robust batch-effect correction framework that preserves radiomic features' variability while reducing technical variations across imaging protocols.

Dimensionality reduction methods were usually applied by minimum absolute shrinkage and selection algorithm (LASSO) in medical data analysis [20]. 10-fold cross-verification was supplied to prevent over-fitting. The hyperparameter "lamda" was optimized though the smallest mean squared error. With only four clinical features available, applying Lasso-based selection risked oversimplification and potential loss of critical clinical information. Only radiomic features and deep radiomic features were involved in feature selection stage in this study. Above feature selection were performed by Python software (version 3.6.4).

Radiomic model construction and validation

To investigate the predictive performance of combined radiomic models, three radiomic models were established, including Model I was generated from traditional radiomic features, Model II was built by traditional radiomic features combined the deep radiomic features, and Model III contained the above two types of radiomic features and clinical features. And then random forest (RF), adaptive boosting (AdaBoost), support vector machine (SVM), K-nearest neighbor (KNN) and Naive Bayes (NB) classifiers were involved to generate these three radiomic models in training cohort. The combined features were used as input into these five classifiers before training. The performance of these models were evaluated by accuracy, sensitivity, specificity, the area under the curve (AUC) and the receiver operating characteristic (ROC) curves in both cohorts.

Statistical analysis

Categorical variables were presented as absolute numbers and percentages, compared by chi-squared test or Fisher exact test. Continuous variables were computed by the mean ± standard deviation, compared by the independent sample t-tests or Wilcoxon tests between cohorts with and without 5-year recurrence. A two-sided p value less than 0.05 was indicated statistically significant. Above statistical analyses were completed in SPSS (version 26.0, IBM Corp, Armonk, NY, USA).

Results

Patients characteristics

In general, this study enrolled 328 LA-NPC patients after NACT two cycles, including 222 cases in the 5-year non-recurrence group and 106 cases in the 5-year recurrence group. The training ($n=229$) and validation cohorts ($n=99$) were divided at in a random ratio of 7:3. The clinical characteristics between two cohorts, with and without recurrence, were found no significant differences (Table 1).

Feature extraction and selection

A total of 975 traditional radiomic features with ICC > 0.75 were respectively extracted from both T1WI+C and T2WI sequences. From above MRI sequences, 1000 deep radiomic features were extracted, respectively.

To quantify the impact of ComBat correction, we performed multivariate analysis of variance (MANOVA) on radiomic features across multi-center. Pre-ComBat analysis revealed significant inter-center variation, indicating strong batch effects. Post-harmonization, MANOVA results showed no significant center-specific differences (see Table S3 for detailed F- and p-values in Supplementary file Sect. 4), confirming successful reduction of technical variability. These findings were further supported by Principal Component Analysis (PCA), which demonstrated that data from different centers clustering relatively more uniformly (Fig. S1-S4 in Supplementary file Sect. 4).

To reduce the dimensionality of radiomic features, LASSO method was performed. The key features were selected with the nonzero coefficients through 1000 iterations. Finally, fifteen radiomic features and six deep radiomic features were chosen from T1WI+C sequence. Nine radiomic features and six deep radiomic features were chosen from T2WI sequence. Selected best lambda parameters were 0.07079 in two types of features of two sequence, respectively. (Fig. 2). The description of these selected features and nonzero coefficients were list in Table 2.

Radiomic model construction and validation

The selected features were used to structure three radiomic models (Model I: traditional radiomic features, Model II: traditional and deep radiomic features, Model III: clinical features and traditional and deep radiomic features). These models were trained with five machine learning classifiers. Predictive performance of these models were shown in Tables 3 and 4. In the validation cohort of T1WI+C sequence, the predictive performance of Model I (AUC range from 0.71 to 0.74) were slightly smaller than Model II

Table 1 Characteristics of LA-NPC patients in the training cohort and validation cohorts

Variable	Training cohort				Validation cohort			
	Non-recurrence ($n=157$)	Recurrence ($n=72$)	X^2/Z	*P*	Non-recurrence ($n=65$)	Recurrence ($n=34$)	X^2/Z	*P*
Gender			0.492	0.483			2.766	0.096
Female	44(28.0)	17(23.6)			17(26.2)	4(11.8)		
Male	113(72.0)	55(76.4)			48(73.8)	30(88.2)		
Age				0.772				0.110
Mean	47.4	46.9			49.5	42.9		
Median	48	47			50	42.5		
Range	14–70	19–72			17–69	11–66		
SD	11.9	11.5			10.2	12.9		
T stage			2.137	0.711			4.228	0.238
I	9(5.7)	2(2.8)			2(3.1)	1(2.9)		
II	8(5.1)	4(5.6)			0(0.0)	1(2.9)		
III	72(45.9)	29(40.3)			36(55.4)	13(38.2)		
IV	68(43.3)	37(51.4)			27(41.5)	19(55.9)		
N stage			5.795	0.122			5.469	0.140
0	15(9.6)	1(1.4)			12(18.5)	1(2.9)		
1	41(26.1)	17(23.6)			14(21.5)	9(26.5)		
2	81(51.6)	42(58.3)			34(52.3)	19(55.9)		
3	20(12.7)	12(16.7)			5(7.7)	5(14.7)		

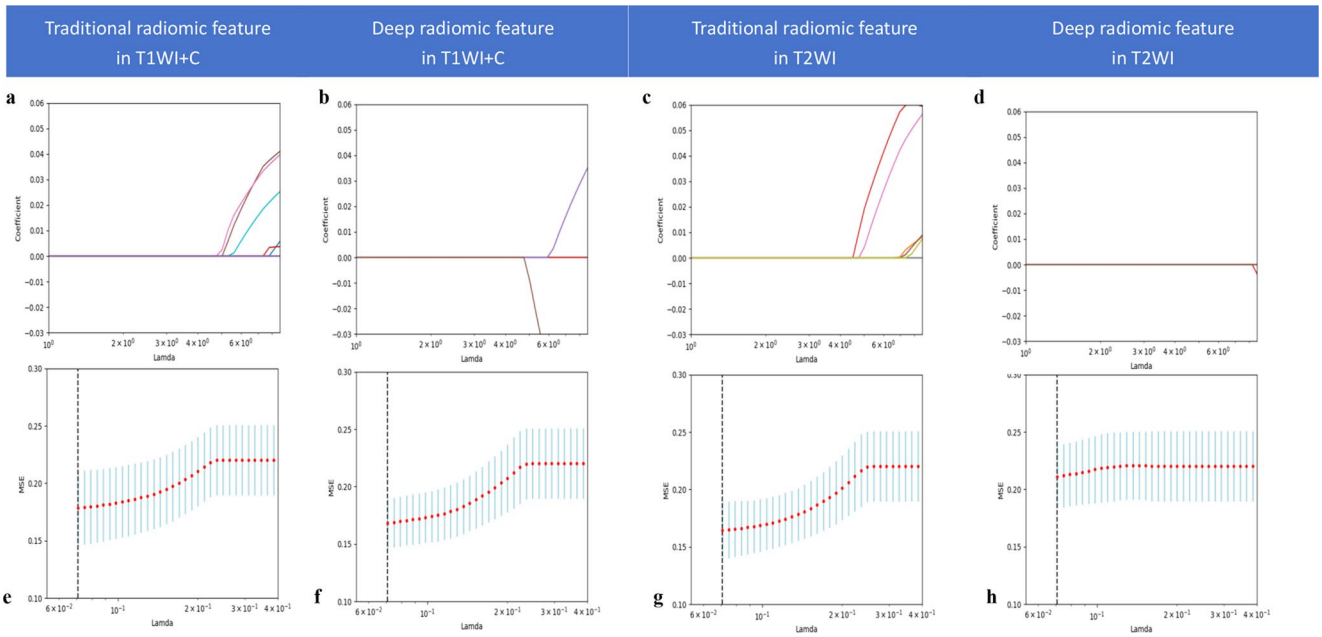

Fig. 2 Multiple radiomic features selection using LASSO in two sequences. First row showed the 10-fold cross-validation curve of handcrafted radiomic features with the optimal lamda value (**a-d**). Second row were the coefficient profiles of radiomic features against the deviance explained (**e-h**)

(AUC range from 0.71 to 0.75) and Model III (AUC range from 0.70 to 0.76). The predictive performance of Model III based on NB was found wonderful in the training cohort [AUC = 0.85 (95% CI: 0.81–0.89)] and in the validation cohort[AUC = 0.76 (95% CI: 0.66–0.85)].

Equally, in T2WI sequence, the predictive performance of Model I (AUC range from 0.80 to 0.84) were less than Model II (AUC range from 0.82 to 0.87) and Model III (AUC range from 0.82 to 0.85). The best performance of the Model II based on RF were showed in the training cohort [AUC = 0.94 (95% CI: 0.92–0.96)] and in the validation cohort [AUC = 0.87 (95% CI: 0.80–0.92)].

Notably, in the validation cohort, the predictive performance of all models in T2WI sequence (AUC range from 0.80 to 0.87) was overall better than that of models in T1WI sequence (AUC range from 0.70 to 0.76). The ROC curve

Table 2 The selected features in the two sequences

Sequence	Feature name	Feature class	Coefficients
T1WI+C	W-LLH_F_E	Firstorder	0.0233
	W-HLL_GLSZM_GLNU	GLSZM	0.0212
	W-HHL_GLSZM_LGZE	GLSZM	-0.0065
	W-LHH_NGTDM_B	NGTDM	0.0019
	W-HLH_F_E	Firstorder	0.0156
	W-HHH_NGTDM_B	NGTDM	0.0482
	W-HLH_GLSZM_GLNU	GLSZM	0.0371
	W-HLH_F_TE	Firstorder	0.0011
	W-LHH_F_E	Firstorder	0.0120
	W-LLH_GLSZM_GLNU	GLSZM	0.0174
	W-LLL_NGTDM_B	NGTDM	0.0183
	G_GLRLM_GLNU	GLRLM	0.0054
	W-HHL_GLRLM_SRHGLE	GLRLM	-0.0099
	W-LLH_F_TE	Firstorder	0.0034
	W-LHH_F_TE	Firstorder	0.0003
	Deep301	Deep radiomic features	-0.0057
	Deep414		-0.0041
	Deep642		-0.0026
	Deep705		0.0217
	Deep909		0.0604
	Deep989		-0.1222
T2WI	L_GLRLM_RV	GLRLM	0.0018
	E_F_TE	Firstorder	0.0222
	W-HLH_GLCM_I	GLCM	0.0088
	W-HLH_NGTDM_B	NGTDM	0.0453
	W-HHH_GLZSM_SZNU	GLZSM	0.0008
	W-HLH_GLZSM_GLNU	GLZSM	0.0272
	O_S_M2DDS	Shape	0.0768
	O_S_M2DDC	Shape	0.0057
	W-HHH_NGTDM_B	NGTDM	0.0396
	Deep66	Deep radiomic features	-0.0143
	Deep104		-0.0071
	Deep129		-0.0048
	Deep154		-0.0459
	Deep770		0.0065
	Deep790		0.0255

Note: B = Busyness; E = Energy; F = Fisrtorder; G = Gradient; I = Imc1;L = Logarithm; O = Original; S = Shape; W = Wavelet; RV = Run-Variance; TE = Total-Energy; HHL = High-High-Low; HLL = High-Low-Low; HHH = High-High-High; LLH = Low-Low-High;LHH = Low-High-High; HLH = High-Low-High; LLL = Low-Low-Low; LGZE = Low-Gray-Level-Zone-Emphasis; GLCM = Gray-Level-Co-occurrence-Matrix; GLNU = Gray-Level-Non-Uniformity; SZNU = Size-Zone-Non-Uniformity; GLSZM = Gray-Level-Size-Zone-Matrix; NGTDM = Neighborhood-Gray-Tone-Difference-Matrix; M2DDC = Maximum-2D-Diameter-Column; GLRLM = Gray-Level-Run-Length-Matrix; GLZSM = Gray-Level-Non-Uniformity; M2DDS = Maximum-2D-Diameter-Slice; SRHGLE = Short-Run-High-Gray-Level-Emphasis

plot of three models based on five classifiers in training and validation cohorts were plot in Figs. 3 and 4.

Discussion

In this study, the traditional and deep radiomic features were extracted from TIWI+C and T2WI sequences, and constructed multiple radiomic models to predict the 5-year recurrence of LA-NPC patients after NACT treatment. Model I (only contained traditional radiomic features), Model II (combined with traditional and deep radiomic features), and Model III (comprised of above two types features and clinical features) were generated based on five machine learning classifiers (SVM, AdaBoost, KNN, NB and RF). In validation cohort of T1WI+C sequence, the Model III based on NB [AUC = 0.76 (95% CI: 0.66–0.85)] showed better predictive performance than other models. In validation cohort of T2WI sequence, the Model II based on RF [AUC = 0.87 (95% CI: 0.80–0.92)] showed better predictive performance than other models. The predictive performance of all models (AUC range from 0.80 to 0.87) in T2WI sequence were better than those models (AUC range from 0.70 to 0.76) in T1WI+C sequence. The predictive performance of Model II was greater than Model I in T2WI sequences. The clinical value of combined radiomic models was obviously improved in assessing the curative effect of the NACT.

In previous studies, traditional radiomic features extracted from MRI images according to the image biomarker standardization initiative (IBSI) [21]. They indicated that these radiomic features are a good indication of the histological

Table 3 The predictive performance of multiple combined radiomic models with T1WI+C sequence in the cohorts

Models in T1WI+C	Training cohort				Validation cohort			
	ACC	SEN	SPE	AUC(95%CI)	ACC	SEN	SPE	AUC(95%CI)
Model I-SVM	0.79	0.47	0.95	0.85(0.80–0.89)	0.76	0.44	0.91	0.73(0.62–0.82)
Model I-RF	0.85	0.64	0.95	0.93(0.90–0.95)	0.75	0.44	0.90	0.71(0.61–0.81)
Model I-KNN	0.82	0.53	0.95	0.87(0.83–0.90)	0.68	0.31	0.85	0.72(0.62–0.81)
Model I-NB	0.77	0.57	0.86	0.82(0.77–0.87)	0.76	0.50	0.88	**0.74**(0.64–0.83)
Model I-Adaboost	0.85	0.73	0.90	0.92(0.89–0.95)	0.75	0.47	0.88	0.72(0.63–0.80)
Model II-SVM	0.84	0.58	0.97	0.88(0.84–0.92)	0.76	0.47	0.90	**0.75**(0.66–0.84)
Model II-RF	0.85	0.66	0.94	0.93(0.90–0.96)	0.77	0.44	0.93	0.74(0.64–0.83)
Model II-KNN	0.86	0.70	0.94	0.94(0.91–0.96)	0.72	0.44	0.85	0.71(0.63–0.79)
Model II-NB	0.80	0.65	0.87	0.85(0.81–0.89)	0.76	0.50	0.88	**0.75**(0.65–0.84)
Model II-Adaboost	0.88	0.74	0.94	0.94(0.91–0.96)	0.74	0.44	0.88	0.73(0.64–0.82)
Model III-SVM	0.81	0.47	0.97	0.89(0.85–0.93)	0.78	0.41	0.96	0.75(0.65–0.84)
Model III-RF	0.86	0.66	0.95	0.94(0.92–0.97)	0.73	0.38	0.90	0.74(0.64–0.83)
Model III-KNN	0.86	0.68	0.94	0.95(0.93–0.96)	0.74	0.41	0.90	0.70(0.60–0.79)
Model III-NB	0.79	0.64	0.86	0.85(0.81–0.89)	0.75	0.47	0.88	**0.76**(0.66–0.85)
Model III-Adaboost	0.88	0.74	0.94	0.94(0.91–0.96)	0.74	0.44	0.88	0.73(0.64–0.82)

SEN, sensitivity; SPE, specificity; ACC, accuracy; AUC, area under the receiver operating characteristic curve; CI, confidence intervals are included in parentheses

SVM, Support vector machine; RF, Random forest; AdaBoost, Adaptive boosting; KNN, K-nearest neighbor; NB, Naive Bayes

Table 4 The predictive performance of multiple combined radiomic models with T2WI sequence in the cohorts

Models in T2WI+C	Training cohort				Validation cohort			
	ACC	SEN	SPE	AUC(95%CI)	ACC	SEN	SPE	AUC(95%CI)
Model I-SVM	0.89	0.69	0.99	0.95(0.92–0.97)	0.76	0.56	0.85	0.82(0.74–0.89)
Model I-RF	0.85	0.65	0.95	0.94(0.92–0.97)	0.75	0.47	0.88	**0.84**(0.77–0.91)
Model I-KNN	0.87	0.73	0.94	0.93(0.91–0.96)	0.78	0.69	0.82	0.80(0.72–0.88)
Model I-NB	0.81	0.62	0.90	0.85(0.81–0.89)	0.76	0.59	0.84	0.80(0.73–0.87)
Model I-Adaboost	0.83	0.66	0.92	0.93(0.90–0.95)	0.76	0.59	0.84	0.82(0.75–0.89)
Model II-SVM	0.93	0.81	0.98	0.98(0.96-1.00)	0.81	0.66	0.88	0.85(0.78–0.91)
Model II-RF	0.86	0.66	0.95	0.94(0.92–0.96)	0.77	0.50	0.90	**0.87**(0.80–0.92)
Model II-KNN	0.83	0.58	0.94	0.89(0.85–0.92)	0.80	0.56	0.91	0.84(0.76–0.91)
Model II-NB	0.79	0.62	0.88	0.85(0.81–0.90)	0.73	0.53	0.82	0.82(0.75–0.88)
Model II-Adaboost	0.86	0.70	0.94	0.94(0.91–0.96)	0.81	0.56	0.93	0.86(0.80–0.92)
Model III-SVM	0.81	0.47	0.97	0.87(0.83–0.91)	0.75	0.41	0.91	0.83(0.77–0.89)
Model III-RF	0.86	0.68	0.95	0.95(0.94–0.97)	0.76	0.47	0.90	**0.85**(0.78–0.92)
Model III-KNN	0.83	0.55	0.97	0.89(0.85–0.93)	0.77	0.44	0.93	0.83(0.76–0.90)
Model III-NB	0.79	0.62	0.88	0.85(0.81–0.90)	0.75	0.56	0.84	0.82(0.75–0.88)
Model III-Adaboost	0.86	0.68	0.94	0.94(0.91–0.96)	0.74	0.47	0.87	**0.85**(0.78–0.91)

SEN, sensitivity; SPE, specificity; ACC, accuracy; AUC, area under the receiver operating characteristic curve; CI, confidence intervals are included in parentheses

SVM, Support vector machine; RF, Random forest; AdaBoost, Adaptive boosting; KNN, K-nearest neighbor; NB, Naive Bayes

heterogeneity of the tumor [22, 23]. In this study, the wavelet filter and the original features were emerged from both sequences, respectively (Table 2). As known, the original features mainly provide the size, shape and texture information of the tumor on the original MRI image. The result in this study showed that the shape feature selected from T2WI sequences were revealed the shape and size information of the tumor. The wavelet filter features reflected the multi-frequency information of the images in the wavelet transform domain [24]. And NGTDM focuses on local intensity differences, making it particularly useful for characterizing subtle texture variations and spatial patterns in medical images [25]. In this study, W-HHH_NGTDM_B were both selected from two sequences, W-LHH_NGTDM_B and W-LLL_NGTDM_B was selected from T1WI+C sequence, W-HLH_NGTDM_B was selected from T2WI sequence, these features were belong to NGTDM class. As mentioned in previous studies, these radiomic features of tumors described pairwise arrangements of pixels with absolute gray difference at a given direction and distance, and are used to emphasize local heterogeneity information from the perspective of grayscale arrangement [11, 24,

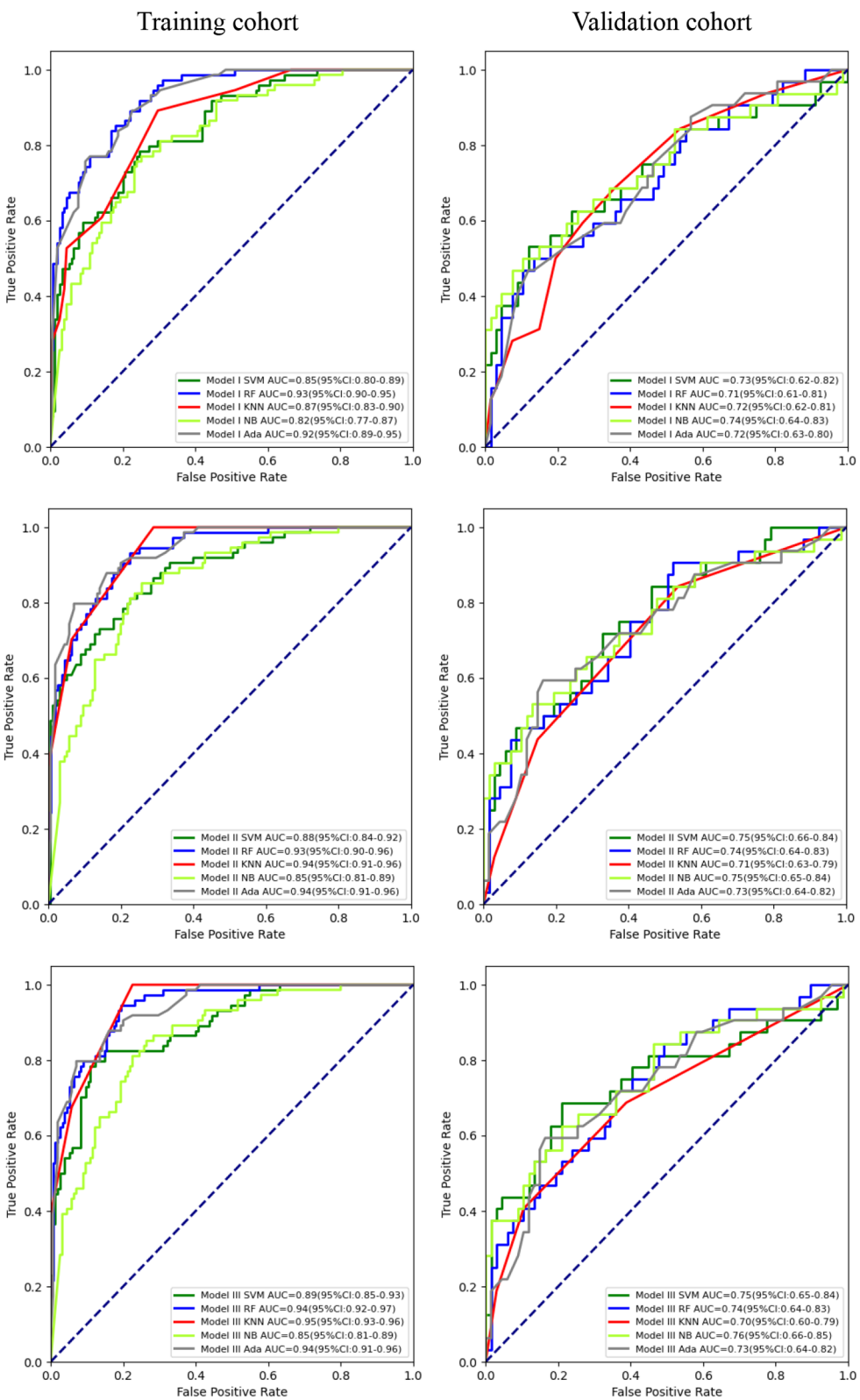

Fig. 3 Predictive performance of multiple radiomic models in T1WI+C sequence. First row showed the predictive performance of the Model I; Second row showed the predictive performance of the Model II; Third row showed the predictive performance of the Model III

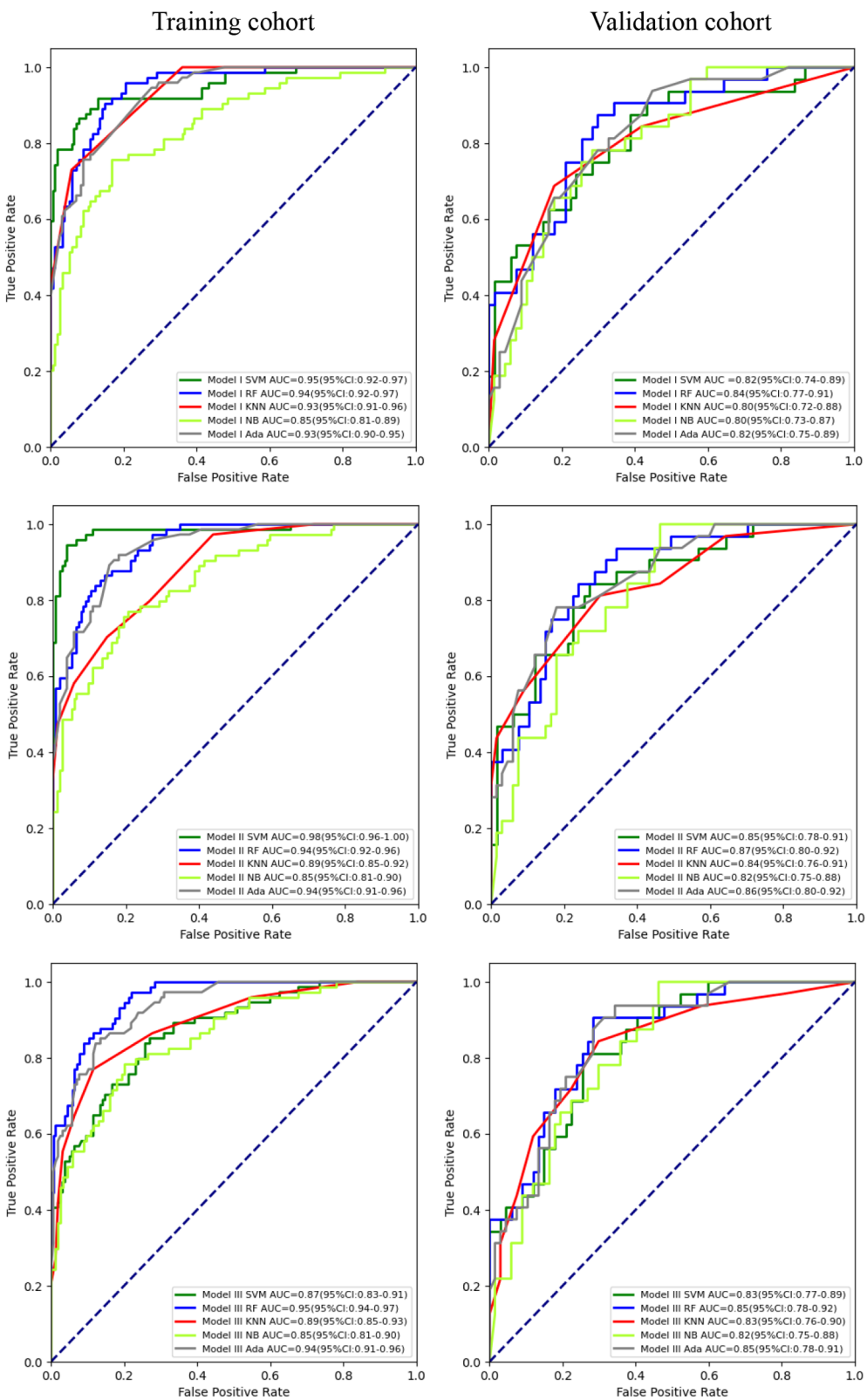

Fig. 4 Predictive performance of multiple radiomic models in T2WI sequence. First row showed the predictive performance of the Model I; Second row showed the predictive performance of the Model II; Third row showed the predictive performance of the Model III

26]. W-HLL_GLSZM_GLNU, W-_LLH_GLSZM_GLNU were selected from T1WI+C sequence. W-HLH_GLSZM_GLNU were selected from two sequences. These belong to GLSZM class. Similar to other studies, GLSZM has potential to distinguish the local complexity texture feature of the tumor [26, 27]. Except above radiomic features, O_S_M2DDS and O_S_M2DDC were selected from T2WI sequence. Visibly, shape features provide quantitative measures of tumor geometry, which can be useful for characterizing lesion morphology, growth patterns, and potential malignancy [25]. These findings indicated that the histological heterogeneity of the tumor was significant, they may be as a meaningful biomarker in predicting the curative effect of the NACT. In this study, the deep radiomic feature could not visually express the shape, size and texture features of the tumor, but the combined models showed the better performance both in two sequences.

To further confirm that combined radiomics models had a better predictive performance, we constructed three models of the two sequences images (T1WI+C and T2WI) using five machine learning classifiers (SVM, RF, KNN, NB, Adaboost). The results showed that Model I performed slightly less than Model II and Model III in two sequences. Especially, in the validation cohort of T2WI sequence, the predictive performance of the Model II based on RF (AUC=0.87) was greater than the Model I based on RF (AUC=0.84). Visibly, combing the traditional radiomic features and deep radiomic features were better predictive power than a single traditional feature. Also it indicated that the deep radiomic feature could be improve the predictive ability. But the predictive performance of all models (AUC range from 0.80 to 0.87) in T2WI sequence were better than those models (AUC range from 0.70 to 0.76) in T1WI+C sequence, shown in Tables 3 and 4. This illustrated that T2WI sequences can help better detect white matter lesions than T1WI+C sequences. As the previous studies reported, a uniform increasing of white matter signal intensity in the T2WI sequence was considered to be indicative of demyelination, gliosis, and edema [27, 28]. Meanwhile, the result suggested that the predictive performance of the Models III were slightly lower than Models II in T2WI sequences. It may be affected by the clinical features. In this study, the clinical features only contained gender, age, T stage, N stage because of the insufficiency of other three outside hospitals of clinical data. And these clinical features were no significant differences were found between cohorts. It may illustrated that these common clinical features could not improve the predictive performance of the models. Although the model demonstrated a relatively good AUC, indicating good overall classification performance, the sensitivity remained notably low(shown in the Tables 3 and 4). This discrepancy may be attributed to the Combat correction process applied to the data. Combat correction is designed to remove batch effects and improve data consistency across different sources or batches. However, it may also inadvertently smooth out subtle biological variations or reduce the discriminative power of certain features that are critical for identifying true positive cases. As a result, while the model's ability to distinguish between classes (as reflected by AUC) remains robust, its capacity to correctly identify positive instances (sensitivity) may be compromised. Future work could explore alternative harmonization methods or feature selection strategies to better preserve biologically relevant signals while mitigating batch effects.

Several limitations still exist. First, although the data were from four hospitals, the sample size from three hospitals might be smaller, in consideration of which we mixed these data to establish a hybrid cohort instead of separated them into an individual one. External validation data are still demanded for further experiment to improve model's generalization. Second, these deep radiomic features could illustrated complete representation of tumor from another angle, but how to visualize the lesion characteristic remains to be studied. Third, our current model utilized gender, age, T stage, and N stage, future studies could improve prediction accuracy by incorporating additional factors such as treatment details (e.g., radiation dose, chemotherapy), tumor biomarkers (e.g., HPV status, EGFR expression), and patient-specific variables (e.g., ECOG performance, comorbidities). Standardized data collection and advanced integration methods (e.g., federated learning) will be essential to achieve this goal. Additionally, our study focused on MRI radiomics due to its superior soft-tissue contrast and clinical utility in NPC, we acknowledge that integrating additional imaging modalities, such as CT and PET, could provide complementary information and potentially improve model performance. Future studies will explore multi-modal imaging analysis to further enhance predictive accuracy.

Conclusion

In brief, we combined traditional radiomic features with deep radiomic features and clinical features to investigate the predictive performance of radiomic models based on machine learning classifiers. These models could be clinically used to assess and predict the curative effect of the LA-NPC patients after NACT treatment. It holds significant implications for advancing the application of deep radiomics models in disease prognosis assessment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10585-025-10349-y.

Acknowledgements None.

Author contributions All authors contributed to the conception and design of the study. Material preparation and data collection were performed by Chunmiao Hu, Qingcheng Meng, ZhianLin, Xinming Huang; data analysis was performed by Li Chen, Congrui Xu. The first draft of the manuscript was written by Chunmiao Hu, Li Chen, Congrui Xu, Jiaxin Chen, Yiling Huang; and all authors commented on previous versions of the manuscript. All authors reviewed the manuscript.

Funding This work was supported by the contract grant from the Natural Science Fund in FuJian Projects (Grant number: 2023J01182), Open Research Fund of Fujian Key Laboratory of Medical Bioinformatics (Grant number: FKLMB-202004).

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval and informed consent The retrospective study was approved by the Fujian Cancer Hospital Ethics Committee and the requirement for informed patient consent was waived.

Competing interests The authors declare no competing interests.

References

1. Bray 1F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Ahmedin Jemal (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3):229–263. https://doi.org/10.3322/caac.21834
2. Huang J, He R, Chen J, Li S, Deng Y, Wu X (2021) Boosting advanced nasopharyngeal carcinoma stage prediction using a Two-Stage classification framework based on deep learning. Int J Comput Intell Syst 14:184. https://doi.org/10.1007/s44196-021-00026-9
3. Weifeng Wang S, Peng H, Wu Y, Luo F, Yuan Z, Lin G, Cheng S, Chen (2021) 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 147(10):2913–2922. https://doi.org/10.1007/s00432-021-03690-8
4. Setakornnukul J, Thephamongkhol K (2018) Neoadjuvant chemotherapy followed by concurrent chemoradiotherapy versus concurrent chemoradiotherapy followed by adjuvant chemotherapy in locally advanced nasopharyngeal carcinoma. BMC Cancer 18(1). https://doi.org/10.1186/s12885-018-4210-3
5. Xixi Liu B, Wu J, Huang Y, Qin Z, Zhang L, Shi X, Ding HQ, Peng G, Kunyu Yang (2022) Tumor factors associated with in-field failure for nasopharyngeal carcinoma after intensity-modulated radiotherapy. Head Neck 44(4):876–888. https://doi.org/10.1002/hed.26983
6. Xue F, Hu C, He X (2017) Long-term patterns of regional failure for nasopharyngeal carcinoma following Intensity-Modulated radiation therapy. J Cancer 8(6):993–999. https://doi.org/10.7150/jca.17858
7. Zhang M-X, Li J, Shen G-P, Zou X, Xu J-J, Jiang R, You R, Hua Y-J, Ma YSJ, Hong M-H, Chen M-Y (2015) Intensity-modulated radiotherapy prolongs the survival of patients with nasopharyngeal carcinoma compared with conventional two-dimensional radiotherapy: A 10-year experience with a large cohort and long follow-up. Eur J Cancer 51(17):2587–2595. https://doi.org/10.1016/j.ejca.2015.08.006
8. Au KH, Roger KC, Ngan, Alice WY, Ng, Darren MC, Poon WT, Ng KT, Yuen, Victor HF, Lee SY, Tung, Anthony TC, Chan, Henry CK, Sze, Ashley CK, Cheng, Anne WM, Lee, Dora LW, Kwong, Anthony HP, Tam (2018) Treatment outcomes of nasopharyngeal carcinoma in modern era after intensity modulated radiotherapy (IMRT) in Hong Kong: A report of 3328 patients (HKNPCSG 1301 study). Oral Oncol 77:16–21. https://doi.org/10.1016/j.oraloncology.2017.12.004
9. Bao D, Zhao Y, Liu Z, Zhong H, Geng Y, Lin M, Li L, Zhao X, Dehong Luo (2021) Prognostic and predictive value of radiomics features at MRI in nasopharyngeal carcinoma. Discover Oncol 12:63. https://doi.org/10.1007/s12672-021-00460-3
10. Zhao L, Gong J, Xi Y, Xu M, Li C, Kang X, Yin Y, Qin W, Yin H, Mei Shi (2020) MRI-based radiomics nomogram May predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma. Eur Radiol 30:537–546. https://doi.org/10.1007/s00330-019-06211-x
11. Bao D, Zhao Y, Li L, Lin M, Zhu Z, Yuan M, Zhong H, Xu H, Zhao X, Dehong Luo (2022) A MRI-based radiomics model predicting radiation-induced Temporal lobe injury in nasopharyngeal carcinoma. Eur Radiol 32:6910–6921. https://doi.org/10.1007/s00330-022-08853-w
12. Xiaohuang Zhuo H, Zhao M, Chen Y, Mu Y, Li J, Cai H, Li Y, Xu Y, Tang (2023) A radiomics model for predicting the response to Methylprednisolone in brain necrosis after radiotherapy for nasopharyngeal carcinoma. Radiat Oncol 18:43. https://doi.org/10.1186/s13014-023-02235-2
13. Wang H, Xu X, Zhang X, Liu Y, Ouyang L, Du P, Li S, Tian Q, Ling J, Guo Y, Lu H (2020) Elaboration of a multisequence MRI-based radiomics signature for the preoperative prediction of the muscle-invasive status of bladder cancer: a double-center study. Eur Radiol 30:4816–4827. https://doi.org/10.1007/s00330-020-06796-8
14. Zheng C, Zhong J, Wang Y, Cao K, Zhang C, Yue P, Xu X, Yang Y, Liu Q, Zou Y, Bingsheng Huang (2024) Deep learning radiomic analysis of MRI combined with clinical characteristics diagnoses placenta accreta spectrum and its subtypes. J Magn Reson Imaging. https://doi.org/10.1002/jmri.29317
15. Yihuai H, Xie C, Yang H, Joshua WK, Ho J, Wen L, Han K-O, Lam, Ian YH, Wong, Simon YK, Law, Keith WH, Chiu, Jianhua Fu (2021) Varut Vardhanabhuti, Computed tomography-based deep-learning prediction of neoadjuvant chemoradiotherapy treatment response in esophageal squamous cell carcinoma. Radiotherapy And Oncology 154:6–13. https://doi.org/10.1016/j.radonc.2020.09.014

16. Paul A, Yushkevich J, Piven HC, Hazlett et al (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 31(3):1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015
17. Joost JM, van Griethuysen A, Fedorov C, Parmar A, Hosny N, Aucoin V, Narayan, Regina GH, Beets-Tan JWL, Aerts (2017) Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 77 (21): e104–e107. https://doi.org/10.1158/0008-5472.CAN-17-0339
18. Shim S-O, Hussain L, Aziz W, Alshdadi AA, Alzahrani A, Omar A (2024) Deep learning convolutional neural network ResNet101 and radiomic features accurately analyzes MpMRI imaging to predict MGMT promoter methylation status with transfer learning approach. Int J Imaging Syst Technol 34(2). https://doi.org/10.1002/ima.23059
19. Kilicarslan G, Koc C, Ozyurt F, Gul Y (2023) Breast lesion classification using features fusion and selection of ensemble ResNet method. Int J Imaging Syst Technol 33(5):1779–1795. https://doi.org/10.1002/ima.22894
20. Lin M, Tang X, Cao L, Liao Y, Zhang Y, Zhou J (2022) Using ultrasound radiomics analysis to diagnose cervical lymph node metastasis in patients with nasopharyngeal carcinoma. Eur Radiol 33:774–783. https://doi.org/10.1007/s00330-022-09122-6
21. Zwanenburg A, Vallières M, Abdalah MA, Hugo JWL, Aerts V, Andrearczyk A, Apte S, Ashrafinia S, Bakas RJ, Beukinga R, Boellaard M, Bogowicz L, Boldrini Irène, Buvat GJR, Cook C, Davatzikos A, Depeursinge M-C, Desseroit N, Dinapoli CV, Dinh S, Echegaray IE, Naqa AY, Fedorov R, Gatta RJ, Gillies V, Goh M, Götz M, Guckenberger SM, Ha M, Hatt F, Isensee P, Lambin S, Leger, Ralph TH, Leijenaar KH, Maier-Hein O, Morin H, Müller S, Napel C, Nioche F, Orlhac S, Pati EAG, Pfaehler JHM, Steenbakkers GC, Troost V, Valentini LV, van Dijk HP, van Velden C, Richter (2020) Steffen Löck The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 295(2), 328–338. https://doi.org/10.1148/radiol.2020191145
22. Marius E, Mayerhoefer A, Materka G, Langs I, Häggström P, Szczypiński P, Gibbs, Cook G (2020) Introduction to radiomics. J Nucl Med 61(4):488–495. https://doi.org/10.2967/jnumed.118.222893
23. Eloyan A, Yue MS, Khachatryan D (2020) Tumor heterogeneity Estimation for radiomics in cancer. Stat Med 39(30):4704–4723. https://doi.org/10.1002/sim.8749
24. Shengping Jiang L, Han L, Liang L Long (2022) Development and validation of an MRI-based radiomic model for predicting overall survival in nasopharyngeal carcinoma patients with local residual tumors after intensity-modulated radiotherapy. BMC Med Imaging 22:174. https://doi.org/10.1186/s12880-022-00902-6
25. Ting-wei F, Malhi H, Varghese B, Cen S, Hwang D, Aron M, Rajarubendra N, Desai M, Vinay Duddalwar (2019) Computed tomography-based texture analysis of bladder cancer: differentiating urothelial carcinoma from micropapillary carcinoma. Abdom Radiol 44:201–208. https://doi.org/10.1007/s00261-018-1694-x
26. Bao D, Liu Z, Geng Y, Li L, Xu H, Zhang Y, Zhao LHX, Zhao Y, Dehong Luo (2022) Baseline MRI-based radiomics model assisted predicting disease progression in nasopharyngeal carcinoma patients with complete response after treatment. Cancer Imaging 22:10. https://doi.org/10.1186/s40644-022-00448-4
27. Zhang B, Lian Z, Zhong L, Zhang X, Dong Y, Chen Q, Zhang L, Mo X, Huang W, Yang W, Zhang S (2020) Machine-learning based MRI radiomics models for early detection of radiation-induced brain injury in nasopharyngeal carcinoma. BMC Cancer 20(1):502. https://doi.org/10.1186/s12885-020-06957-4
28. Huang L, Yang Z, Zeng Z, Ren H, Jiang M, Hu Y, Xu Y, Zhang H, Ma K, Long L (2023) MRI-based radiomics models for the early prediction of radiation-induced Temporal lobe injury in nasopharyngeal carcinoma. Front Neurol 14. https://doi.org/10.3389/fneur.2023.1135978