

RESEARCH ARTICLE

Open Access



# MR radiomics predicts pathological complete response of esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy: a multicenter study

Yunsong Liu<sup>1</sup>, Yi Wang<sup>2</sup>, Xin Wang<sup>1</sup>, Liyan Xue<sup>3</sup>, Huan Zhang<sup>2</sup>, Zeliang Ma<sup>1</sup>, Heping Deng<sup>4</sup>, Zhaoyang Yang<sup>3</sup>, Xujie Sun<sup>3</sup>, Yu Men<sup>5</sup>, Feng Ye<sup>6</sup>, Kuo Men<sup>1</sup>, Jianjun Qin<sup>7</sup>, Nan Bi<sup>1</sup>, Qifeng Wang<sup>2\*</sup> and Zhonguang Hui<sup>5\*</sup>

## Abstract

**Background** More than 40% of patients with resectable esophageal squamous cell cancer (ESCC) achieve pathological complete response (pCR) after neoadjuvant chemoradiotherapy (nCRT), who have favorable prognosis and may benefit from an organ-preservation strategy. Our study aims to develop and validate a machine learning model based on MR radiomics to accurately predict the pCR of ESCC patients after nCRT.

**Methods** In this retrospective multicenter study, eligible patients with ESCC who underwent baseline MR (T2-weighted imaging) and nCRT plus surgery were enrolled between September 2014 and September 2022 at institution 1 (training set) and between December 2017 and August 2021 at institution 2 (testing set). Models were constructed using machine learning algorithms based on clinical factors and MR radiomics to predict pCR after nCRT. The area under the curve (AUC) and cutoff analysis were used to evaluate model performance.

**Results** A total of 155 patients were enrolled in this study, 82 in the training set and 73 in the testing set. The radiomics model was constructed based on two radiomics features, achieving AUCs of 0.968 (95%CI 0.933–0.992) in the training set and 0.885 (95%CI 0.800–0.958) in the testing set. The cutoff analysis resulted in an accuracy of 82.2% (95%CI 72.6–90.4%), a sensitivity of 75.0% (95%CI 58.3–91.7%), and a specificity of 85.7% (95%CI 75.5–96.0%) in the testing set.

**Conclusion** A machine learning model based on MR radiomics was developed and validated to accurately predict pCR after nCRT in patients with ESCC.

**Keywords** Esophageal neoplasms, Magnetic resonance, Treatment outcome, Neoadjuvant Chemoradiotherapy

\*Correspondence:  
Qifeng Wang  
littlecancer@163.com  
Zhonguang Hui  
drhuizg@163.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Neoadjuvant chemoradiotherapy (nCRT) combined with surgery has successfully improved the survival of patients with resectable esophageal cancer and has become a standard treatment [1]. Two large randomized control trials, CROSS and NEOCRTEC5010, showed that 43.2–49% of patients with esophageal squamous cell carcinoma (ESCC) were confirmed as pathological complete response (pCR) in surgical resection specimens after nCRT [2, 3], which is related to favorable prognosis [4]. Esophagectomy is known to have a high incidence of surgical complications and a significant decrease in the quality of life [5, 6]. Active surveillance is an organ-preserving strategy in which patients predicted to reach pCR after nCRT continue surveillance rather than surgery. Researches indicate it achieves similar disease control and survival compared to esophagectomy after nCRT [7]. The success of this strategy hinges on accurately predicting pCR after nCRT to select suitable patients who can benefit from organ preservation and to ensure the efficacy of treatment in patients with residual disease. Therefore, prediction of pCR is important for the individualized treatment of ESCC.

Previous studies have explored the prediction of nCRT response for esophageal cancer by conventional evaluation methods based on CT, PET/CT, or endoscopic biopsy. However, the accuracy of these methods is limited because of the difficulty in distinguishing tumor tissues from reactive changes [8]. Radiomics, which extracts numerous features from images, has shown promising results in predicting pCR in esophageal cancer after nCRT based on CT or PET/CT, with reported areas under the curve (AUCs) of 0.65–0.85 in testing sets [9–13].

MR has high tissue resolution, contains a tremendous amount of information, and has been improved to have superior sensitivity relative to CT and PET in diagnostic performance for esophageal cancer [14]. Conventional MR image analysis has shown promising potential in predicting pCR after nCRT in esophageal cancer. A meta-analysis including seven studies with 158 patients showed that the increase of apparent diffusion coefficient (ADC) computed from diffusion-weighted imaging (DWI) during nCRT was significantly different between pCR and non-pCR groups [15]. However, no study has explored MR radiomics, which can maximize the use of valuable image information, to predict nCRT response in esophageal cancer. In this context, we aimed to develop an accurate early prediction model using a machine learning method based on MR radiomics to predict pCR after nCRT in ESCC.

## Methods

### Patients

The institutional review boards approved this retrospective study of the two institutions, and the requirement for informed consent was waived.

Consecutive patients with ESCC were included between September 2014 and September 2022 at Institution 1 (National Cancer Center, Beijing, China) and between December 2017 and August 2021 at Institution 2 (Sichuan Cancer Hospital & Institution, Sichuan, China). The inclusion criteria were as follows: (1) histologically confirmed ESCC; (2) underwent neoadjuvant concurrent chemoradiation followed by radical esophagectomy; (3) age 18–80 years; (4) Karnofsky performance status  $\geq 70$ ; and (5) underwent pretreatment fat-suppressed T2-weighted imaging (T2WI). Patients were excluded if they had any of the following conditions: (1) distant metastasis (except for supraclavicular lymph node metastasis) or (2) insufficient T2WI quality (with MR artifacts existing inside the primary tumor).

Pretreatment staging started with an in-depth medical history and thorough physical examination. Patients then underwent a series of diagnostic evaluations, including routine hematologic and biochemical tests, enhanced CT scans of the neck, thorax, and abdomen, upper gastrointestinal endoscopy with biopsy, and endoscopic ultrasonography. Cervical ultrasonography, augmented by fine-needle aspiration, was employed when lymph node involvement was suspected. Additional diagnostic procedures, such as bronchoscopy, positron emission tomography (PET) and radionuclide bone imaging, were optional if clinically indicated. The details of the neoadjuvant chemoradiotherapy regimens are provided in eAppendix 1. Before surgery, clinical restaging was performed through upper gastrointestinal endoscopy with endoscopic ultrasonography and neck-thorax-abdomen CT. PET was indicated if distant progression was suspected. Patients at institutions 1 and 2 were allocated to the training and testing sets, respectively.

### Pathological assessment

Two senior pathologists (L. X. and Z. Y.) specializing in esophageal cancer with more than 10 years of experience performed pathological assessments. Pathological assessment included detailed evaluations of the tumor, such as its location, type, and histological grade, along with the depth of tumor invasion and the status of resection margins. The tumor's response to treatment was assessed using the Mandard tumor regression grade (TRG) system. Additionally, lymph node status was thoroughly examined, including the location, number of nodes affected, and the extent of therapeutic response in these nodes. The pCR was defined as the absence of residual

tumor cells in all resected specimens, including the primary tumor site and lymph nodes.

### MRI protocol

Patients received baseline MR scan in 2 weeks before nCRT. Fat-suppressed T2WI were accessed from two institutions. At institution 1, all MR examinations were performed using a GE Discovery MR750w 3.0T scanner or a GE Discovery MR750 3.0T scanner. At institution 2, MR examinations were performed using four scanners (SIEMENS Skyra 3.0T scanner, SIEMENS Avanto 1.5T scanner, UIH uMR780 3.0T scanner, and UIH uMR588 1.5T scanner). Details of all scanner protocols are provided in eTable 1.

### Segmentation of regions of interest (ROIs)

The 3D-ROIs of the primary tumor were manually segmented on axial T2WI independently at the two institutions. Images of patients in the training set were segmented using ITK-SNAP software by one radiologist (Y. L.) with 4 years of experience and reviewed by a senior radiologist (Z. H.) with more than 20 years of experience who was blinded to the pathological information. Images of patients in the testing set were segmented using MIM software by one radiologist (H. Z.) with four years of experience and reviewed by a senior radiologist (Y. W.) with 14 years of experience in a blinded manner. Disagreements were resolved through discussion until a consensus was reached. The segmentation of 10 randomly selected patients from the training set was performed by another radiologist (Z. M.) with five years of experience. Inter-class correlation coefficients (ICCs) were calculated using the two sets of segmentation in the training set and

features with ICCs greater than 0.75 were considered satisfactory reproducibility for further analysis.

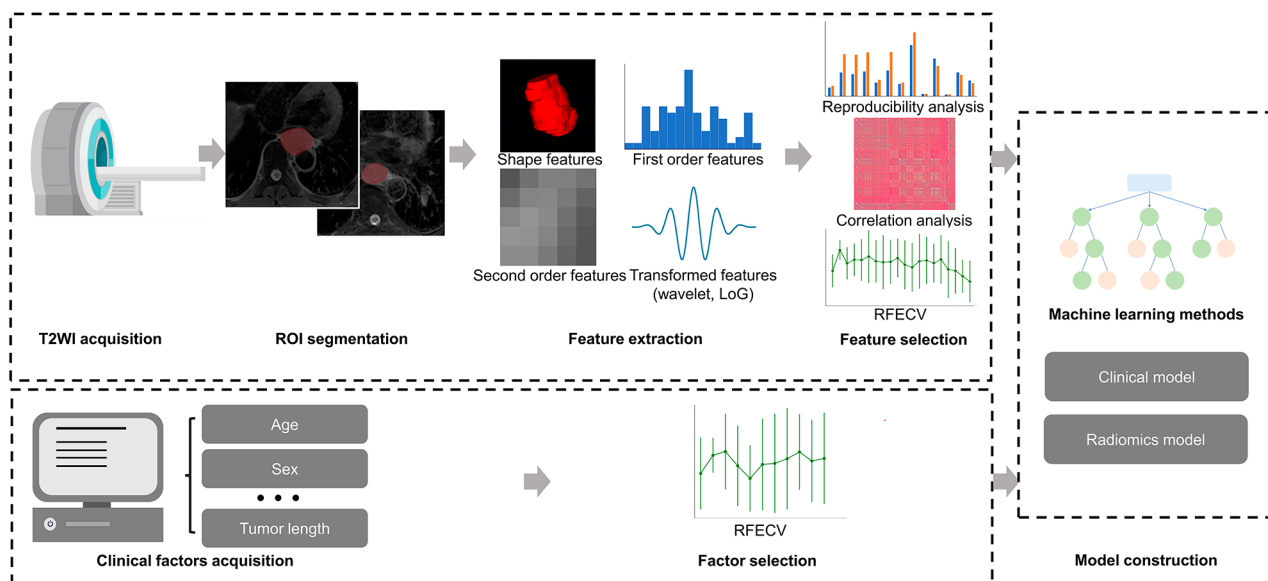
### Radiomics feature extraction

The radiomics features from ROIs were computed using the PyRadiomics package (version 3.0.1), as recommended by the IBSI [16]. First, the z-score method was used to normalize the distribution of image densities. The voxel size was resampled to 1 mm × 1 mm × 1 mm. Features were extracted from the original, wavelet-filtered, and Laplacian of Gaussian-filtered images. Features with satisfactory reproducibility were harmonized in the two institutions using the ComBat method to compensate for multivendor effects [17]. Further details on the definitions and algorithms of radiomics features are provided in eAppendix 2.

### Feature selection and modeling

Feature selection was performed independently in the training set. Clinical factors were selected using recursive feature elimination with 5-fold cross-validation (RFECV). The RFECV method used backward elimination to iteratively remove the feature with the least contribution to the predictive performance of the classifier. It generated rankings of features based on the number of iterations when the feature was removed. This process continued until the model performance became worse in the cross-validation, leaving the most important features.

Radiomics features with satisfactory reproducibility were selected using Pearson correlation analysis ( $|r|$  threshold of 0.95) to eliminate redundancy. RFECV was then performed to select the optimal features from the remaining features for radiomics model construction.



**Fig. 1** Analysis flowchart. T2WI, T2-weighted imaging; ROI, region of interest; RFECV, recursive feature elimination with cross validation

Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest and XGBoost were adopted as the machine learning algorithms for model construction on the training set. The hyperparameters were tuned using the grid-search method in the training set. Internal five-fold cross-validation were performed in the training set to generate mean AUCs. The constructed models were applied independently to the testing set. The workflow of model development is shown in Fig. 1.

### Statistical analysis

The receiver operating characteristic (ROC) curve and AUC were used to evaluate the model performance. The cutoffs were determined using the Youden index in the training set. They applied to the testing set to calculate the accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). We utilized 1000 bootstraps to generate a 95% confidence interval (95%CI). A decision curve analysis was performed to evaluate the benefits of the models in clinical applications. The DeLong and integrated discrimination improvement (IDI) tests were performed to compare the predictive performances of different models.

Statistical analyses were performed using SPSS version 26.0, Python version 3.9, and R version 4.1.2. Categorical variables were compared using Fisher's exact test, and continuous variables were compared using the Mann-Whitney U test. Kaplan-Meier analysis was utilized to evaluate the disease-free survival (DFS) and log-rank tests were performed between pCR group and non-pCR group, in actual population and predicted population, respectively. Statistical tests were two-tailed, and  $P < 0.05$  was considered significant. The data processing and model construction code is publicly available ([https://github.com/NCCYUNSONG/ESO\\_MR\\_RESEARCH.git](https://github.com/NCCYUNSONG/ESO_MR_RESEARCH.git)).

## Results

### Patient characteristics

A total of 655 patients were screened, 155 of whom were included in this study. A flowchart of patient selection is shown in Fig. 2. Eighty-two patients (66 men and 16 women, median age 62 years [IQR 55–66]) and 73 patients (61 men and 12 women, median age 62 years [IQR 55–67]) were eligible for the training (institution 1) and testing sets (institution 2), respectively. The median radiation doses were 41.4 Gy (IQR 37.8–41.4) in the training set and 40.0 Gy (IQR 40.0–40.0) in the testing set. Sixty-three (76.8%) patients in the training set received simultaneous integrated boost of PGTV (median dose 47.5 Gy, IQR [44.9–49.2]), which was not applied to patients in the testing

set. The intervals between nCRT and surgery were 57 days (IQR 48–79) and 47 days (IQR 40–52) in the training and testing sets, respectively. Thirty-nine (47.5%) patients achieved pCR in the training set, and 24 (32.9%) patients achieved pCR in the testing set. The patients' clinical characteristics are presented in Table 1.

There were no statistically significant differences in the clinical characteristics between pCR and non-pCR patients in the training and testing sets, except for sex ( $P = 0.03$ ) and tumor location ( $P = 0.045$ ) in the testing set (eTable 2).

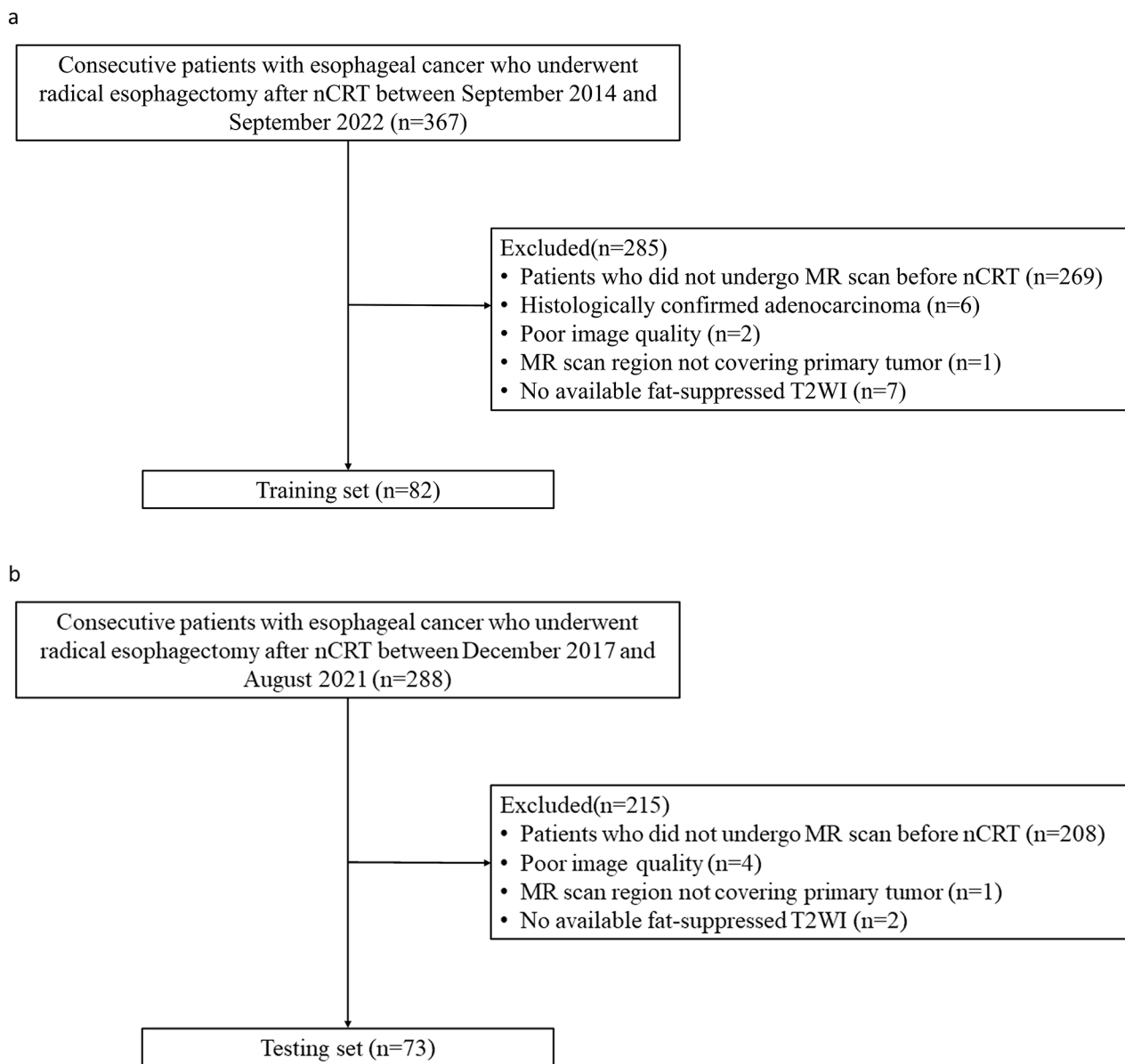
### Performance of the clinical model

The candidates for clinical factors were the same as those listed in Table 1. Age, tumor length, and interval between nCRT and surgery were included in the model after filtration (eFigure 1a). The best model using logistic regression (eTable 3) achieved AUCs in the training and testing set being 0.592 (95%CI 0.472–0.716) and 0.584 (95%CI 0.441–0.714), respectively (Fig. 3a; Table 2). The accuracies were 58.5% (95%CI 53.7–70.7%) and 61.6% (95%CI 43.8–75.3%) in the training and testing sets, respectively. Internal cross-validation showed a mean AUC of 0.596 (eFigure 2a).

### MR radiomic analysis

In total, 1106 features were extracted from each patient. The details of the extracted features are listed in eTable 4. Satisfactory reproducibility was achieved in 928 (83.9%) of the 1106 features with an ICC threshold of 0.75. After correlation analysis with a Pearson  $|r| > 0.95$  to remove multicollinear features, 299 features remained. The top two features, which were both wavelet features, of the ranking by RFECV were finally selected to build the radiomics model (eFigure 1b, eTable 4).

The best radiomics model was constructed using Random Forest (eTable 3), with AUCs of 0.968 (95%CI 0.933–0.992) in the training set and 0.885 (95%CI 0.800–0.958) in the testing set, respectively (Fig. 4a; Table 2). Accuracies were 92.7% (95%CI 87.8–97.6%) and 82.2% (95%CI 72.6–90.4%), respectively. In the training set, the sensitivity, specificity, PPV, and NPV were 87.2% (95%CI 76.9–97.4%), 97.7% (95%CI 93.0–100%), 97.1% (95%CI 91.4–100%), and 89.4% (95%CI 82.4–97.6%), respectively. In the testing set, the sensitivity, specificity, PPV and NPV were 75.0% (95%CI 58.3–91.7%), 85.7% (95%CI 75.5–96.0%), 72.0% (95%CI 58.6–89.5%), and 87.5% (95%CI 80.0–95.5%), respectively (Table 2). Internal cross-validation showed a mean AUC of 0.805 (eFigure 2b). Satisfactory calibration was achieved (Fig. 4b), and decision curve analysis confirmed the clinical benefit of the radiomics model (eFigure 3). The predictive performance of the



**Fig. 2** **a**, Selection process of institution 1 (training set); **b**, Selection process of institution 2 (testing set). nCRT, neoadjuvant chemoradiotherapy; T2WI, T2-weighted imaging

radiomics model in both training and testing set was significantly higher than those of the clinical model using the DeLong ( $P < 0.001$  in training and testing sets) and IDI ( $P < 0.001$  in the training set and  $P = 0.001$  in the testing set) tests (Table 3).

The analysis of radiomics features across various scanning parameters revealed that the selected features demonstrated commendable stability, exhibiting no statistically significant differences when subjected to multiple parameters (eTable 6). Furthermore, the radiomics model exhibited robust performance across distinct scanning parameters, achieving AUCs of 1.000 and 0.971 in the training set, and 0.891, 1.000, and

0.833 in the testing set, for different scanning parameters (eFigure 5).

Correlation analysis revealed that the two chosen radiomics features had a significant association with clinical T stage ( $P = 0.031$ ) and a marginally significant relationship with clinical N stage ( $P = 0.067$ ), respectively (eFigure 5).

With a median follow-up of 32 months (IQR 16–46) in the training set and 24 months (IQR 16–32) in the testing set, the radiomics model's predicted pCR group showed significantly longer DFS than the predicted non-pCR group in both training and testing sets, consistent with the actual pCR and non-pCR groups (all  $P < 0.05$ ) (eFigure 6).

**Table 1** Patient Characteristics

Characters	Institution 1 (training set) N=82	Institution 2 (testing set) N=73	P-value
pCR	39 (47.5)	24 (32.9)	0.073
Age, years (median, IQR)	62 (56, 66)	62 (55, 67)	0.836
Sex			0.304
Female	16 (19.6)	11 (15.1)	
Male	66 (80.4)	62 (84.9)	
KPS			<0.001
≤80	48 (58.5)	13 (17.8)	
>80	34 (41.5)	60 (82.2)	
Location			0.935
Neck	0 (0.0)	1 (1.4)	
Upper thoracic	9 (11.0)	7 (9.6)	
Middle thoracic	29 (35.4)	28 (38.3)	
Lower thoracic	42 (51.2)	36 (49.3)	
Gastroesophageal junction	2 (2.4)	1 (1.4)	
Length, cm (median, IQR)	5 (4.0, 6.2)	5 (4.0, 7.0)	0.461
cT			0.781
1	1 (1.2)	1 (1.4)	
2	4 (4.9)	3 (4.1)	
3	53 (64.6)	53 (72.6)	
4	24 (29.3)	16 (21.9)	
cN			0.173
0	9 (11.0)	2 (2.7)	
1	22 (26.8)	27 (37.0)	
2	42 (51.2)	36 (49.3)	
3	9 (11.0)	8 (11.0)	
cM			0.030
0	76 (92.7)	73 (100)	
1	6 (7.3)	0 (0.0)	
Chemotherapy regimen			0.009
Platinum based	62 (75.6)	66 (90.4)	
Non-platinum based	20 (24.4)	7 (9.6)	
Radiotherapy technique			<0.001
IMRT	13 (15.9)	73 (100)	
VMAT	69 (84.1)	0 (0.0)	
Radiation dose, Gy (median, IQR)	41.4 (37.8, 41.4)	40.0 (40.0, 40.0)	0.142
Simultaneous integrated boost	63 (76.8)	0 (0.0)	<0.001
Interval between nCRT and surgery, days (median, IQR)	57 (48, 79)	47 (40, 52)	<0.001

Data are presented as n (%) unless otherwise stated. pCR, pathological complete response; IQR, interquartile range; KPS, Karnofsky Performance Status; cT, clinical T stage; cN, clinical N stage; cM, clinical M stage; IMRT, Intensity-Modulated Radiation Therapy; VMAT, Volumetric Modulated Arc Therapy; nCRT, neoadjuvant chemoradiotherapy

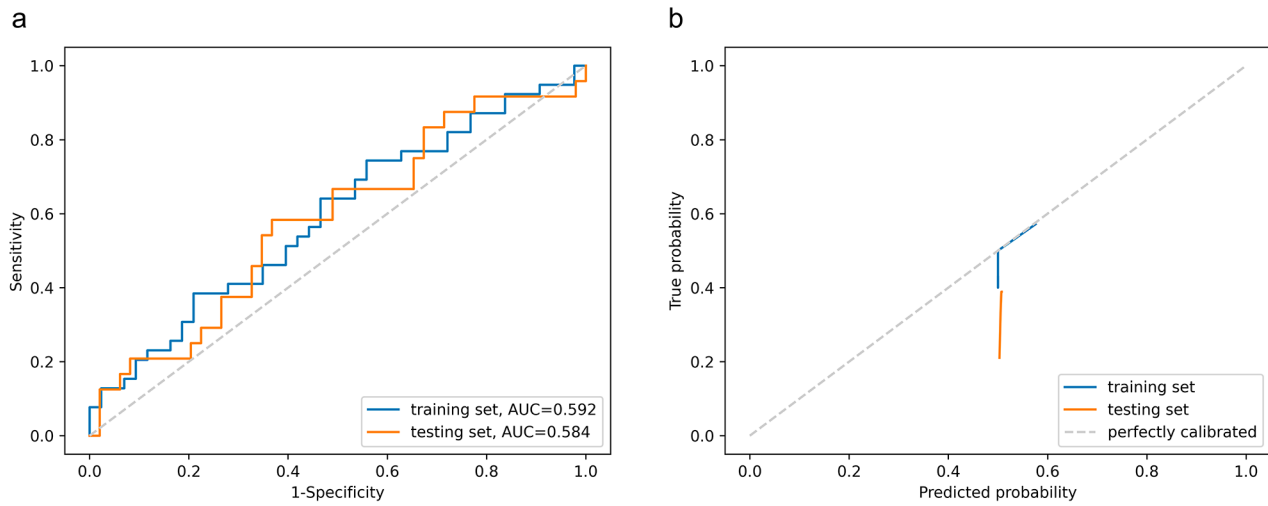
## Discussion

A machine learning model based on MR radiomics was developed to predict pCR in patients with ESCC after nCRT precisely. The model demonstrated satisfactory predictive performance in the external testing set with high AUC, sensitivity, and specificity, which can assist in the implementation of individualized treatment for

ESCC. To the best of our knowledge, this is the first study to use MR radiomics and the first to evaluate the predictive value of MR in an external testing set, in predicting pCR in esophageal cancer after nCRT.

In this study, we established a radiomics model using pretreatment T2WI of patients with ESCC after nCRT to predict pCR, with an AUC of 0.885 (95%CI 0.800–0.958) in the external testing set. The performance of the radiomics model was significantly better than that of the clinical model. Previous studies have also shown that clinical factors cannot accurately predict pCR after nCRT for esophageal cancer [18, 19] and are generally not included in the final models of radiomics studies [31]. Similarly, the clinical model established using demographic, tumor, and treatment-related features showed poor predictive ability (AUC 0.584, 95%CI 0.441–0.714).

The radiomics model developed in this study outperforms those reported in similar studies. Previous studies focusing on the value of radiomics have primarily used CT and PET/CT to predict pCR after nCRT for esophageal cancer. For example, van Rossum et al. [13] used PET images from 217 patients with esophageal cancer to extract radiomic features before and after nCRT, achieving an AUC of 0.77 in internal testing. Yang et al. [12] used pretreatment CT images to build a radiomics model to predict pCR, achieving an AUC of 0.79 in the testing set. However, these studies lacked external testing, limiting their results' generalizability. Hu et al. [9] conducted a multicenter study, building a radiomics model with CT images of 161 patients in the training set to predict pCR in an external testing set of 70 patients with an AUC of 0.852. However, imaging protocols were limited, which could affect their generalizability. MR has high soft tissue resolution and abundant signal information. A meta-analysis showed that the pooled sensitivity and specificity of MR in predicting pCR for esophageal cancer after nCRT were 80% and 83%, respectively. In contrast, those of CT were 35% and 83%, and those of PET were 62% and 73%, respectively [8]. Li et al. [20] built radiomics models for predicting the treatment response in patients with colorectal cancer after neoadjuvant chemotherapy, achieving an AUC of 0.766 for CT and 0.859 for T2WI in the testing set. Fat-suppressed T2WI has the advantages of delineating the tumor from surrounding fat and providing better information about the tumor's extension, which is indicated to lead to better evaluation of tumors compared to conventional T2WI [21]. Hou et al. [22] compared radiomics based on fat-suppressed T2WI and conventional T2WI in predicting response to nCRT according to RECIST in ESCC and found fat-suppressed T2WI leading to higher accuracy. In our study, the MR radiomics-based model based on fat-suppressed T2WI achieved accurate prediction (AUC=0.885) in the external testing set of 73 patients with images obtained

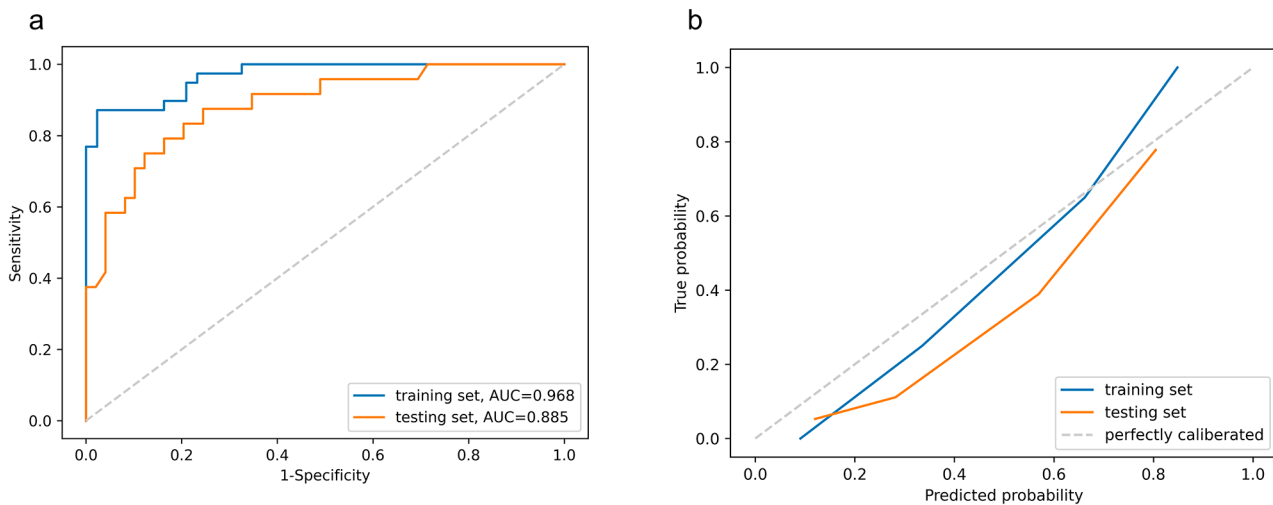


**Fig. 3** Performance of the clinical model. **a**, Receiver operating characteristic curves; **b**, Calibration curves. AUC, area under the curve

**Table 2** Predictive performance of final models

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
<b>Clinical model</b>						
Training set	0.592 (0.472–0.716)	58.5 (48/82) [53.7–70.7]	74.4 (29/39) [12.8–94.9]	44.1 (19/43) [20.9–100]	54.7 (29/53) [51.0–100]	65.5 (19/29) [55.3–85.7]
Testing set	0.584 (0.441–0.714)	61.6 (45/73) [43.8–75.3]	58.3 (14/24) [16.7–100]	63.3 (31/49) [20.4–98.0]	43.8 (14/32) [36.1–87.5]	75.6 (31/41) [70.3–100]
<b>Radiomics model</b>						
Training set	0.968 (0.933–0.992)	92.7 (76/82) [87.8–97.6]	87.2 (34/39) [76.9–97.4]	97.7 (42/43) [93.0–100.0]	97.1 (34/35) [91.4–100.0]	89.4 (42/47) [82.4–97.6]
Testing set	0.885 (0.800–0.958)	82.2 (60/73) [72.6–90.4]	75.0 (18/24) [58.3–91.7]	85.7 (42/49) [75.5–96.0]	72.0 (18/25) [58.6–89.5]	87.5 (42/48) [80.0–95.5]

Data in parentheses are numerators and denominators, with 95% CIs in brackets. AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value



**Fig. 4** Performance of the radiomics model. **a**, Receiver operating characteristic curves; **b**, Calibration curves. AUC, area under the curve

from four different MR scanners, and satisfactory performances were achieved separately for different scanners, demonstrating excellent predictive performance and

generalizability compared to previous studies, which is potentially valuable for clinical application. Additionally, the predicted pCR group exhibited superior DFS relative

**Table 3** Comparison of predictive performance of models

Model comparison	DeLong test P-value	IDI test P-value
Radiomics model (Training set) vs. Clinical model (Training set)	< 0.001	< 0.001
Radiomics model (Testing set) vs. Clinical model (Testing set)	< 0.001	0.001

IDI, integrated discrimination improvement

to the non-pCR group, highlighting the prognostic value of our model. This reinforces its potential as an effective aid in clinical decision-making processes.

Our model expands the domain of predictive models and demonstrates better accuracy, although a few studies have shown promising potential for MR-based prediction of neoadjuvant therapy efficacy in esophageal cancer. In a prospective study, the ADC was used to predict pCR after nCRT for esophageal cancer, achieving an AUC of 0.791 [23]. Meta-analysis revealed that changes of ADC during treatment significantly associated with pCR after neoadjuvant therapy for esophageal cancer [15]. However, baseline ADC was not associated with pathological response, indicating the ADC may not be competent for early prediction [15]. Besides, the lack of testing sets in these studies limits their reliability. Vollenbrock et al. [24] investigated the accuracy of qualitative assessment using T2WI and DWI, with an AUC of 0.7. Nevertheless, although it is valuable to predict pCR based on DWI, small sample size, lacking of external validation and various cut-offs of ADC in those studies limit its clinical applicability and the precision is highly dependent on multiple time-point imaging. In contrast, our model used pretreatment images to perform an early prediction with high accuracy in both training and testing sets of relatively larger sample sizes. The selected features are associated with clinical T stage and N stage, suggesting that these features may reflect the state of disease progression to some extent. Qu et al. [25] conducted a single-center study using dynamic contrast-enhanced MRI to extract features and predict responders (Mandard TRG 1–2) after neoadjuvant chemotherapy for esophageal cancer, achieving an AUC of 0.86 (95% CI 0.74–0.97) in the testing set. However, neoadjuvant chemotherapy is only recommended for locally advanced esophageal adenocarcinoma [26]. Taken together, these studies and our current research support the value of MR in predicting the efficacy of neoadjuvant therapy for esophageal cancer.

The present study has several limitations. First, the sample size of the training set was relatively small. Future studies with larger sample sizes may improve the model performance. Second, all the patients included in our study had ESCC, and the predictive performance for the adenocarcinoma was unknown. Third, this was a retrospective study conducted at two centers, and the generalizability of the results requires further validation.

Fourth, due to the retrospective nature of this study, MR sequences of patients were imbalanced and some valuable sequences such as DWI were not performed in majority of patients. However, building on the findings of our current study, we are going to undertake a prospective study that employs predefined sequences including DWI with various b-values, as well as dynamic contrast-enhanced MR, and the scanning will be conducted at multiple time-points, to enhance the precision and performance of our model further. Finally, with the increasing research on neoadjuvant immunotherapy for esophageal cancer, it is necessary to conduct large-scale prospective multicenter studies incorporating multiple MR sequences, molecular biomarkers, and gene expression data to establish accurate and generalizable predictive models.

## Conclusions

In conclusion, our study established and validated a machine learning model based on MR radiomics to predict pCR after nCRT for ESCC, providing a helpful tool for response prediction and may potentially assist personalized treatment. Further validation using larger independent datasets is necessary.

## Abbreviations

nCRT	Neoadjuvant chemoradiotherapy
ESCC	Esophageal squamous cell carcinoma
pCR	Pathological complete response
AUC	Area under the curve
ADC	Apparent diffusion coefficient
DWI	Diffusion-weighted imaging
T2WI	T2-weighted imaging
ROI	Region of interest
RFECV	Recursive feature elimination with 5-fold cross validation
ROC	Receiver operating characteristics
PPV	Positive predictive value
NPV	Negative predictive value
IDI	Integrated discrimination improvement
IQR	Interquartile range

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40644-024-00659-x>.

Supplementary Material 1

Supplementary Material 2

## Acknowledgements

Not applicable.

## Authors' contributions

All authors contributed to the study conception and design. Y. Liu, Y. Wang, X. Wang, and L. Xue contributed equally to the research. Y. Liu: resources, data curation, formal analysis, methodology, writing - original draft; Y. Wang: resources, data curation, writing- review&editing; X. Wang: resources, data curation, writing- review&editing; L. Xue: resources, data curation, writing- review&editing; H. Zhang: visualization, writing-original draft; Z. Ma: methodology, writing-original draft; H. Deng: resources, methodology; Z. Zhao: resources, data curation; X. Sun: data curation, writing- review&editing; Y. Men: data curation, writing- review&editing; F. Ye: resources, investigation; K. Men: methodology, software; J. Qin: resources, investigation, writing- review&editing; N. Bi: supervision, investigation; Q.



Wang: project administration; supervision; writing- review&editing; Z. Hui: project administration; supervision; writing- review&editing. All authors read and approved the final manuscript.

#### Funding

This study has received funding by Beijing Hope Run Special Fund of Cancer Foundation of China (LC2022R03).

#### Data availability

The data used and analyzed during the current study are available from the corresponding authors on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

Institutional Review Board approval was obtained in National Cancer Center, Beijing, China, and Sichuan Cancer Hospital & Institution, Sichuan, China. Written informed consent was waived.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

##### Author details

<sup>1</sup>Department of Radiation Oncology, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Panjiayuan Nanli #17, Chaoyang District, Beijing, China

<sup>2</sup>Department of Radiation Oncology, Radiation Oncology Key Laboratory of Sichuan Province, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, No.55, Section 4, South Renmin Road, Chengdu 610042, China

<sup>3</sup>Department of Pathology, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Panjiayuan Nanli #17, Chaoyang District, Beijing, China

<sup>4</sup>Department of Diagnostic Radiology, Radiation Oncology Key Laboratory of Sichuan Province, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, No.55, Section 4, South Renmin Road, Chengdu, China

<sup>5</sup>Department of VIP Medical Services & Radiation Oncology, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Panjiayuan Nanli #17, Chaoyang District, Beijing 100021, China

<sup>6</sup>Department of Diagnostic Radiology, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Panjiayuan Nanli #17, Chaoyang District, Beijing, China

<sup>7</sup>Department of Thoracic Surgery, National Clinical Research Center for Cancer/Cancer Hospital, National Cancer Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Panjiayuan Nanli #17, Chaoyang District, Beijing, China

Received: 29 October 2023 / Accepted: 10 January 2024

Published online: 23 January 2024

#### References

- Shapiro J, van Lanschot JJB, Hulshof MCCM, van Hagen P, van Berge Henegouwen MI, Wijnhoven BPL, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol.* 2015;16(9):1090–8.
- van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge Henegouwen MI, Wijnhoven BP, et al. Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med.* 2012;366(22):2074–84.
- Yang H, Liu H, Chen Y, Zhu C, Fang W, Yu Z, et al. Neoadjuvant Chemoradiotherapy followed by surgery versus surgery alone for locally advanced squamous cell carcinoma of the Esophagus (NEOCRTEC5010): a phase III Multicenter, Randomized, open-label clinical trial. *J Clin Oncol.* 2018;36(27):2796–803.
- Berger AC, Farma J, Scott WJ, Freedman G, Weiner L, Cheng JD, et al. Complete response to neoadjuvant chemoradiotherapy in esophageal carcinoma is associated with significantly improved survival. *J Clin Oncol.* 2005;23(19):4330–7.
- Schandl A, Lagergren J, Johar A, Lagergren P. Health-related quality of life 10 years after oesophageal cancer surgery. *Eur J Cancer.* 2016;69:43–50.
- Low DE, Kuppasamy MK, Alderson D, Cecconello I, Chang AC, Darling G, et al. Benchmarking complications Associated with Esophagectomy. *Ann Surg.* 2019;269(2):291–8.
- van der Wilk BJ, Eyck BM, Hofstetter WL, Ajani JA, Piessen G, Castoro C, et al. Chemoradiotherapy followed by active Surveillance Versus Standard Esophagectomy for Esophageal Cancer: a systematic review and individual Patient Data Meta-analysis. *Ann Surg.* 2022;275(3):467–76.
- de Gouw DJJM, Klarenbeek BR, Driessen M, Bouwense SAW, van Workum F, Fütterer JJ, et al. Detecting pathological complete response in Esophageal Cancer after Neoadjuvant Therapy based on imaging techniques: a diagnostic systematic review and Meta-analysis. *J Thorac Oncol.* 2019;14(7):1156–71.
- Hu Y, Xie C, Yang H, Ho JWK, Wen J, Han L, et al. Assessment of Intratumoral and Peritumoral Computed Tomography Radiomics for Predicting Pathological Complete response to Neoadjuvant Chemoradiation in patients with esophageal squamous cell carcinoma. *JAMA Netw Open.* 2020;3(9):e2015927.
- Beukinga RJ, Hulshoff JB, Mul VEM, Noordzij W, Kats-Ugurlu G, Slart R, et al. Prediction of response to Neoadjuvant Chemotherapy and Radiation Therapy with Baseline and Restaging (18)F-FDG PET imaging biomarkers in patients with esophageal Cancer. *Radiology.* 2018;287(3):983–92.
- Rishi A, Zhang GG, Yuan Z, Sim AJ, Song EY, Moros EG, et al. Pretreatment CT and (18) F-FDG PET-based radiomic model predicting pathological complete response and loco-regional control following neoadjuvant chemoradiation in oesophageal cancer. *J Med Imaging Radiat Oncol.* 2021;65(1):102–11.
- Yang Z, He B, Zhuang X, Gao X, Wang D, Li M, et al. CT-based radiomic signatures for prediction of pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant chemoradiotherapy. *J Radiat Res.* 2019;60(4):538–45.
- van Rossum PSN, Fried DV, Zhang L, Hofstetter WL, van Vulpen M, Meijer GJ, et al. The incremental value of subjective and quantitative Assessment of 18F-FDG PET for the prediction of Pathologic Complete Response to Preoperative Chemoradiotherapy in Esophageal Cancer. *J Nucl Med.* 2016;57(5):691–700.
- Lee SL, Yadav P, Starekova J, Christensen L, Chandereeng T, Chappell R, et al. Diagnostic performance of MRI for Esophageal Carcinoma: a systematic review and Meta-analysis. *Radiology.* 2021;299(3):583–94.
- Maffazzoli L, Zilio MB, Klamt AL, Duarte JA, Mazzini GS, Campos VJ, et al. ADC as a predictor of pathologic response to neoadjuvant therapy in esophageal cancer: a systematic review and meta-analysis. *Eur Radiol.* 2020;30(7):3934–42.
- Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: standardized quantitative Radiomics for High-Throughput Image-based phenotyping. *Radiology.* 2020;295(2):328–38.
- Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate Multicenter effects affecting CT Radiomics. *Radiology.* 2019;291(1):53–9.
- MacGuill M, Mulligan E, Ravi N, Rowley S, Byrne PJ, Hollywood D, et al. Clinicopathologic factors predicting complete pathological response to neoadjuvant chemoradiotherapy in esophageal cancer. *Dis Esophagus.* 2006;19(4):273–6.
- Molena D, Sun HH, Badr AS, Mungo B, Sarkaria IS, Adusumilli PS, et al. Clinical tools do not predict pathological complete response in patients with esophageal squamous cell cancer treated with definitive chemoradiotherapy. *Dis Esophagus.* 2014;27(4):355–9.
- Li ZY, Wang XD, Li M, Liu XJ, Ye Z, Song B, et al. Multi-modal radiomics model to predict treatment response to neoadjuvant chemotherapy for locally advanced rectal cancer. *World J Gastroenterol.* 2020;26(19):2388–402.

21. Lu DS, Saini S, Hahn PF, Goldberg M, Lee MJ, Weissleder R, et al. T2-weighted MR imaging of the upper part of the abdomen: should fat suppression be used routinely? *AJR Am J Roentgenol.* 1994;162(5):1095–100.
22. Hou Z, Li S, Ren W, Liu J, Yan J, Wan S. Radiomic analysis in T2W and SPAIR T2W MRI: predict treatment response to chemoradiotherapy in esophageal squamous cell carcinoma. *J Thorac Dis.* 2018;10(4):2256–67.
23. Li QW, Qiu B, Wang B, Wang DL, Yin SH, Yang H et al. Prediction of pathologic responders to neoadjuvant chemoradiotherapy by diffusion-weighted magnetic resonance imaging in locally advanced esophageal squamous cell carcinoma: a prospective study. *Dis Esophagus.* 2018;31(2).
24. Vollenbrock SE, Voncken FEM, van Dieren JM, Lambregts DMJ, Maas M, Meijer GJ, et al. Diagnostic performance of MRI for assessment of response to neoadjuvant chemoradiotherapy in oesophageal cancer. *Br J Surg.* 2019;106(5):596–605.
25. Qu J, Ma L, Lu Y, Wang Z, Guo J, Zhang H, et al. DCE-MRI radiomics nomogram can predict response to neoadjuvant chemotherapy in esophageal cancer. *Discov Oncol.* 2022;13(1):3.
26. Ajani JA, D'Amico TA, Bentrem DJ, Cooke D, Corvera C, Das P, et al. Esophageal and Esophagogastric Junction Cancers, Version 2.2023, NCCN Clinical Practice guidelines in Oncology. *J Natl Compr Canc Netw.* 2023;21(4):393–422.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.