ONCOLOGY

Multiparametric magnetic resonance imaging-derived radiomics for the prediction of disease-free survival in early-stage squamous cervical cancer

Yan Zhou[1] · Hai-Lei Gu[2] · Xin-Lu Zhang[2] · Zhong-Fu Tian[2] · Xiao-Quan Xu[1] · Wen-Wei Tang[2]

Abstract

Objective To conduct multiparametric magnetic resonance imaging (MRI)-derived radiomics based on multi-scale tumor region for predicting disease-free survival (DFS) in early-stage squamous cervical cancer (ESSCC).

Methods A total of 191 ESSCC patients (training cohort, $n = 135$; validation cohort, $n = 56$) from March 2016 to September 2019 were retrospectively recruited. Radiomics features were derived from the T2-weighted imaging (T2WI), contrast-enhanced T1-weighted imaging (CET1WI), diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC) map for each patient. DFS-related radiomics features were selected in 3 target tumor volumes (VOI^{entire}, $VOI^{+5\,mm}$, and $VOI^{-5\,mm}$) to build 3 rad-scores using the least absolute shrinkage and selection operator (LASSO) Cox regression analysis. Logistic regression was applied to build combined model incorporating rad-scores with clinical risk factors and compared with clinical model alone. Kaplan–Meier analysis was used to further validate prognostic value of selected clinical and radiomics characteristics.

Results Three radiomics scores all showed favorable performances in DFS prediction. Rad-score ($VOI^{+5\,mm}$) performed best with a C-index of 0.750 in the training set and 0.839 in the validation set. Combined model was constructed by incorporating age categorized by 55, Federation of Gynecology and Obstetrics (Figo) stage, and lymphovascular space invasion with rad-score ($VOI^{+5\,mm}$). Combined model performed better than clinical model in DFS prediction in both the training set (C-index 0.815 vs 0.709; $p = 0.024$) and the validation set (C-index 0.866 vs 0.719; $p = 0.001$).

Conclusion Multiparametric MRI-derived radiomics based on multi-scale tumor region can aid in the prediction of DFS for ESSCC patients, thereby facilitating clinical decision-making.

Key Points
- *Three radiomics scores based on multi-scale tumor region all showed favorable performances in DFS prediction. Rad-score ($VOI^{+5\,mm}$) performed best with favorable C-index values.*
- *Combined model incorporating multiparametric MRI-based radiomics with clinical risk factors performed significantly better in DFS prediction than the clinical model.*
- *Combined model presented as a nomogram can be easily used to predict survival, thereby facilitating clinical decision-making.*

Keywords Cervical cancer · Magnetic resonance imaging · Radiomics · Disease-free survival

---

Yan Zhou and Hai-Lei Gu contribute equally to this work

✉ Xiao-Quan Xu
  xiaoquanxu_1987@163.com

✉ Wen-Wei Tang
  tww3077@163.com

[1] Department of Radiology, The First Affiliated Hospital of Nanjing Medical University, Gulou District, No. 300, Guangzhou Rd, Nanjing 210029, People's Republic of China

[2] Department of Radiology, Women's Hospital of Nanjing Medical University, No. 123, Mochou Rd, Qinhuai District, Nanjing 210029, People's Republic of China

Abbreviations

ADC	Apparent diffusion coefficient
BIC	Bayesian information criterion
CET1WI	Contrast-enhanced T1-weighted imaging
CI	Confidence interval
DFS	Disease-free survival
DWI	Diffusion-weighted imaging
ESSCC	Early-stage squamous cervical cancer
FIGO	Federation of Gynecology and Obstetrics
HPV	Human papillomavirus
ICC	Intraclass correlation coefficient
LASSO	Least absolute shrinkage and selection operator
LVI	Lymphovascular space invasion
MRI	Magnetic resonance imaging
SCC	Squamous cervical cancer
SCCA	Squamous cell carcinoma antigen
T2WI	T2-weighted imaging
t-ROC	Time-dependent receiver operating characteristic
VIF	Variance inflation factor
VOIs	Volumes of interest

Introduction

Cervical cancer is the fourth most common cause of cancer-related mortality in women worldwide, and squamous cervical cancer (SCC) accounts for 75–80% of cases [1]. With the popularity of cancer screening, the incidence of early-stage (IB-IIA) SCC (ESSCC) has increased dramatically during the past decades [1, 2]. Radical hysterectomy with bilateral pelvic lymph node dissection is the cornerstone of treatment for ESSCC [3]. Despite curative treatment, 25–30% of patients with ESSCC experience locoregional recurrence or distant metastasis [2]. Therefore, accurate prognosis prediction is essential for choosing the optimal surgical plan, postoperative adjuvant therapy, and the intensity of postoperative observation.

Federation of Gynecology and Obstetrics (Figo) staging is most commonly used to evaluate cervical cancer prognosis in clinical practice [1, 4]. However, Figo staging is easily affected by physician's experience, and it is limited by the lack of tumor's pathological characteristics. Some pathological characteristics such as tumor differentiation, lymph node metastasis, and lymphovascular space invasion (LVI) have been identified as adverse prognostic factors for cervical cancer in previous studies [5, 6]. Therefore, the new 2018 Figo staging incorporates the pathological finding into the disease staging [4]. However, the pathological information can only be obtained through invasive methods, and it is easily influenced by sampling error and interobserver variability. Therefore, there is a need for noninvasive prognostic indicators that provide quantifiable risk measures for patients with cervical cancer.

Given its excellent contrast resolution, magnetic resonance imaging (MRI) is the first-line imaging modality to diagnose, stage, and monitor cervical cancer [7, 8]. In addition to detailed anatomical information, MRI can describe the tumor microscopic characteristics through some functional MRI sequences, such as contrast-enhanced T1-weighted imaging (CET1WI) and diffusion-weighted imaging (DWI). A meta-analysis has revealed that DWI and its derived apparent diffusion coefficient (ADC) may be a suitable candidate indicator of recurrence in patients with cervical cancer [9]. However, MRI-derived quantitative metrics are usually expressed as average values, which ignore the heterogeneity of tumors.

Radiomics is a mathematical-statistical procedure that quantifies high-throughput features from medical images, and enables noninvasive profiling of tumor heterogeneity [10, 11]. In the field of cervical cancer, previous studies have verified that radiomics might be a promising tool in predicting tumor staging, LVI, and lymph node status [12–15]. Fang et al reported an MRI-derived radiomics score as a prognostic biomarker in patients with cervical cancer [16]; however, only T2-weighted imaging (T2WI) and CET1WI were used, while functional DWI and ADC maps were not incorporated. Moreover, previous studies have usually performed the extraction of radiomics features based on the entire volumetric tumor region, but peritumoral information was ignored. Yet, some previous studies have shown the significant application of peritumoral radiomics in tumor assessment [17, 18].

Therefore, this study aimed to conduct multiparametric MRI-derived radiomics based on multi-scale tumor region for predicting disease-free survival (DFS) in ESSCC patients.

Materials and methods

Study population

Our institutional review board approved this retrospective study and written informed consent was not required. A total of 346 consecutive patients with suspicious cervical cancer for preoperative pelvic MRI assessment from March 2016 to September 2019 were retrospectively enrolled. The inclusion criteria were as follows: (1) patients underwent radical hysterectomy and bilateral pelvic lymph node dissection; (2) postoperative pathology confirmed to be ESSCC (Figo stage IB-IIA); (3) patients had no history of chemotherapy or radiotherapy before MRI examination; (4) image quality was adequate for analysis; (5) clinical data were complete.

Table 1 Characteristics of patients with early-stage cervical cancer in the training and validation sets

Characteristics	Training set (135)	Validation set (56)	p value
Age (years, mean ± SD)	50.3 ± 10.0	49.9 ± 9.2	0.978
Age (years)			0.794
≤ 50	72 (53.3%)	28 (50.0%)	
> 50	63 (46.7%)	28 (50.0%)	
Age (years)			0.999
≤ 55	84 (62.2%)	35 (62.5%)	
> 55	51 (37.8%)	21 (37.5%)	
Erythrocytes (10^{-9}/L, mean ± SD)	4.2 ± 0.5	4.3 ± 0.5	0.103
Neutrophils (10^{-9}/L, mean ± SD)	6.4 ± 2.0	6.5 ± 2.8	0.686
Platelets (10^{-9}/L, mean ± SD)	236.9 ± 69.0	237.0 ± 69.0	0.904
HPV			0.880
Negative	66 (48.9%)	26 (46.4%)	
Positive	69 (51.1%)	30 (53.6%)	
CEA (ng/mL, mean ± SD)	4.4 ± 7.8	3.7 ± 7.2	0.360
CA125 (u/mL, mean ± SD)	21.1 ± 23.9	17.9 ± 13.4	0.806
CA199 (u/mL, mean ± SD)	16.0 ± 24.0	12.7 ± 8.6	0.943
SCCA (ng/mL, mean ± SD)	8.8 ± 12.8	9.3 ± 11.6	0.485
Size (cm, mean ± SD)	3.7 ± 1.1	3.5 ± 1.3	0.379
Figo stage			0.231
IB	65 (48.1%)	33 (58.9%)	
IIA	70 (51.9%)	23 (41.1%)	
Differentiation			0.837
Low grade	47 (34.8%)	21 (37.5%)	
Middle grade	75 (55.6%)	31 (55.4%)	
High grade	13 (9.6%)	4 (7.1%)	
Lymph node metastasis			0.568
Non-metastasis	94 (69.6%)	42 (75.0%)	
Metastasis	41 (30.4%)	14 (25.0%)	
Number of lymph node metastasis			0.797
≤ 2	117 (86.7%)	50 (89.3%)	
> 2	18 (13.3%)	6 (10.7%)	
LVI			0.093
Negative	37 (27.4%)	23 (41.1%)	
Positive	98 (72.6%)	33 (58.9%)	
Invasion depth			0.181
≤ 2/3	38 (28.1%)	22 (39.3%)	
> 2/3	97 (71.9%)	34 (60.7%)	
Additional chemotherapy			0.825
Yes	22 (16.3%)	10 (17.9%)	
No	113 (83.7%)	46 (82.1%)	
Rad-score (VOI^{entire}) (medians ± IQ)	− 0.7 ± 3.4	− 0.7 ± 3.1	0.879
Rad-score (VOI$^{+5\,mm}$) (medians ± IQ)	− 0.7 ± 4.1	− 0.6 ± 5.6	0.769
Rad-score (VOI$^{-5\,mm}$) (medians ± IQ)	− 1.1 ± 3.0	− 1.0 ± 2.8	0.798
DFS time (months, mean ± SD)	29.1 ± 13.6	27.2 ± 13.4	0.424

Data are numbers of patients and parentheses indicate the proportion if not specified

HPV, human papillomavirus; *SCCA*, squamous cell carcinoma antigen; *Figo*, Federation of Gynecology and Obstetrics; *LVI*, lymphovascular space invasion; *DFS*, disease-free survival; *IQ*, interquartile

Finally, 191 patients (mean age, 50 years; range, 24–73 years) were included in the study. The patients were allocated to a training set and a validation set on basis of the time of surgery at 7:3 ratio. The training set consisted of 135 patients (mean age, 50 years; range, 24–73 years) from March 2016 to August 2018, whereas the validation set consisted of 56 patients (mean age, 50 years; range, 32–69 years) from September 2018 to September 2019. Detailed demographic characteristics are summarized in Table 1.

Postoperative follow-up

All of the patients underwent radical hysterectomy and bilateral pelvic lymph node dissection. Thirty-two patients (training set, $n=22$; validation set, $n=10$) received adjuvant chemoradiotherapy postoperatively. The patients were routinely followed up every 3–6 months during the first 2 years, 6 months for the next 3–5 years, and then once a year thereafter. Disease-free survival (DFS) was defined as the period from the date of surgery to the date of first locoregional recurrence, distant metastasis, or last follow-up. Locoregional recurrence and distant metastasis were screened by means of gynecological examination and cross-sectional imaging such as computed tomography (CT), MRI, and positron emission tomography CT, or proven by biopsy [1, 2].

MRI acquisitions and parameters

All of the patients were examined using 1.5-T MRI (Achieva, Philips Healthcare) using a 16-channel matrix torso coil. For contrast scanning, a gadolinium contrast (Omniscan, GE HealthCare) was intravenously administered at 0.1 mL/kg at 2.5 mL/s, followed by 20 mL of saline flushing. Detailed MRI scan parameters are listed in Table S1.

Clinical and pathological information

Clinical information including age, blood routine examination (erythrocytes, neutrophils, platelets), human papillomavirus (HPV) status, and serum gynecological tumor markers (CEA, CA125, CA199, and squamous cell carcinoma antigen (SCCA)) was obtained from medical records [1]. Age was categorized by 50 and 55 (years) separately to obtain a binary variable. Tumor size was determined as the longest diameter of the lesion with the maximum tumor size. Pathological information included tumor differentiation, lymph node metastasis, LVI, and invasion depth. The number of lymph node metastases was categorized by two. Figo staging was determined following the 2018 Figo staging guidelines [4].
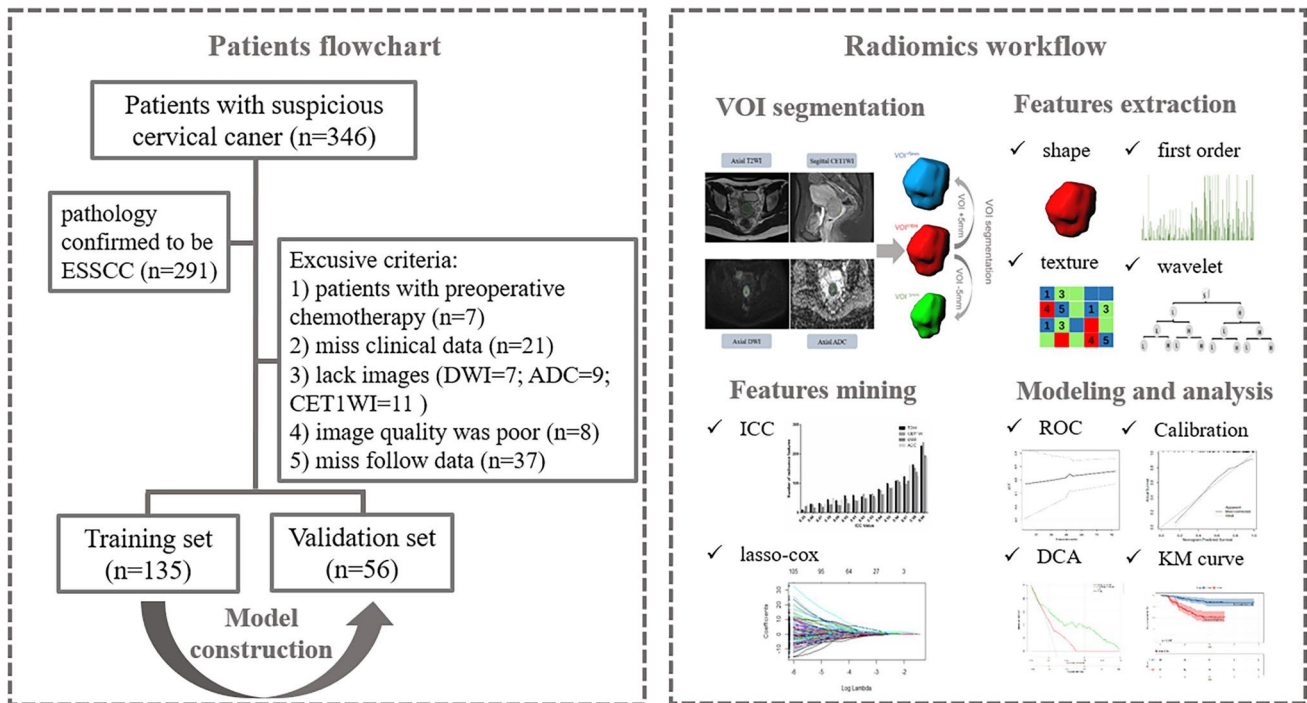

Fig. 1 The study flowchart and the radiomics workflow

Multi-scale tumor segmentation and radiomics feature extraction

Tumor segmentation was conducted by reader 1 (with 6 years of experience in gynecological MRI) and confirmed again by reader 2 (with 8 years of experience in gynecological MRI). Both readers were blinded to clinical information and pathological results.

The workflow of our study is shown in Fig. 1. Tumor segmentation was performed semi-automatically using syngo.via Frontier Radiomics (Siemens Healthcare) on axial oblique T2WI, DWI, ADC, and sagittal CET1WI [19]. The software allowed semi-automatic entire volumetric interest (VOI^{entire}) segmentation through integrated hybrid algorithm combining a threshold-based approach with model-based morphological processing [19, 20]. The method consisted of three parts: (1) identification of points on the tumor boundary by region growing and subsequent ray casting from the seed point; (2) calculation of an ellipsoid that approximates the shape of the tumor; (3) convex hull of the inverse mask within the dilated ellipsoid [20]. Then, regions at a 5-mm distance outer (VOI$^{+5\,mm}$) and inner (VOI$^{-5\,mm}$) from the tumor surface were automatically reconstructed using specified dilate and erode tools, which applied the morphological operations of erosion and dilation to the target VOI^{entire}. Finally, three types of VOIs were generated (VOI^{entire}, VOI$^{+5\,mm}$, and VOI$^{-5\,mm}$), and the contour of each VOI was locally refined mainly in the involved tumors with obscure boundary. The process of multi-scale tumor segmentation is depicted in Fig. 2.
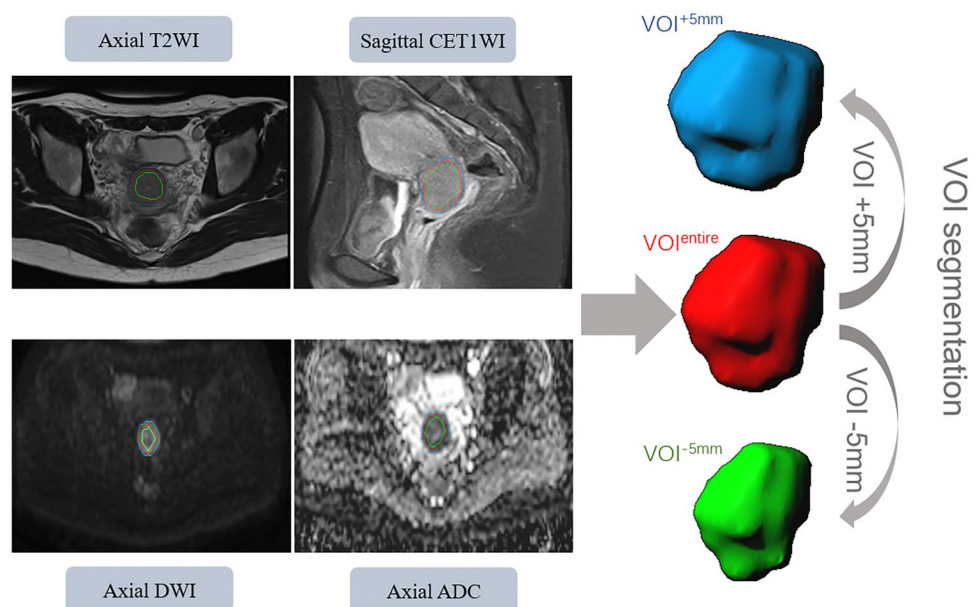
After tumor segmentation, radiomics features from each VOI were automatically computed using syngo.via Frontier Radiomics interfaces with the PyRadiomics library [21]. Radiomics features, including 17 shape-based features, 18

first-order statistical features, and 75 textural features, were extracted from the original image set. Textural features included five classes (24 Gray Level Co-occurrence Matrix (GLCM), 14 Gray Level Dependence Matrix (GLDM), 16 Gray Level Size Zone Matrix (GLSZM), 16 Gray Level Run Length Matrix (GLRLM), 5 Neighbouring Gray Tone Difference Matrix (NGTDM)). Detailed information on the texture features is listed in Table S2. Default parameter settings are available from the website (https://github.com/AIM-Harvard/pyradiomics/tree/master/examples/exampleSettings). To obtain high-throughput features, the non-linear intensity transforms on image voxels (square, square root, logarithm, and exponential); Laplacian of Gaussian (LoG) filtering with sigma values of 0.5, 1.5, 2.5, 3.5, and 4.5 mm; and eight wavelet transformations algorithms (LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH) were conducted for first-order statistical and textural features (Table S3). Standardization with z scores was performed for all extracted radiomics features to eliminate the impact of the different ranges of gray values.

Radiomics feature selection and three radiomics scores' calculation

A two-step procedure was applied for high-dimensional radiomics feature selection. First, 75 patients were randomly selected for test and re-test and intraclass correlation coefficient (ICC) was calculated to evaluate the reproducibility of the features. Features with ICC higher than 0.8 were remained in the subsequent analysis. Second, least absolute shrinkage and selection operator (LASSO) Cox regression was applied for radiomics feature selection and radiomics

Fig. 2 VOI segmentation for the radiomics analysis. A radiologist first segmented the whole tumor region (VOI^{entire}) semi-automatically with the help of the software (red line). On basis of VOI^{entire} of the tumor, the regions at 5 mm outer (VOI$^{+5\,mm}$, blue line) and inner (VOI$^{-5\,mm}$, green line) from the tumor boundary surface were automatically reconstructed, respectively, using the object operations module integrated in the software

Axial T2WI

Sagittal CET1WI

Axial DWI

Axial ADC

VOI^{+5mm}

VOI^{entire}

VOI^{-5mm}

VOI +5mm

VOI -5mm

VOI segmentation

scores' (rad-scores) construction. LASSO Cox regression compressed the coefficients of useless features to zero with the regulation parameter λ through tenfold cross-validation [16]. Rad-scores were then constructed by a linear combination of final selected robust features with their respective coefficients.

On account of three different types of VOIs, three radiomics scores (rad-score (VOI$^{\text{entire}}$), rad-score (VOI$^{+5\text{ mm}}$), and rad-score (VOI$^{-5\text{ mm}}$)) were finally built through the above-mentioned procedure, respectively.

Establishment, performance, and validation of clinical and combined models

Together with the three constructed radiomics scores, clinical predictors of recurrence with statistical significance in the univariable Cox regression analysis were included in multivariate Cox regression analysis. Backward stepwise variable selection was implemented with the Bayesian information criterion (BIC) [22]. Variance inflation factor (VIF) was checked for each variable included in the final combined model [23]. The clinical model, including only the clinical risk factors, was also built for comparison.

DFS probabilities were estimated using the Kaplan–Meier method and compared using the log-rank test [24]. Prognostic predictive ability of the models was assessed and compared using C-index calculated as prognostic performance indicator [25]. Time-dependent receiver operating characteristic (t-ROC) analysis with the areas under the curve at different time points was applied to assess prognostic accuracy [26]. Calibration curves were used to evaluate calibration precision. Internal validation was tested in an independent validation cohort.

Clinical utility

The clinical usefulness of the constructed models was evaluated by decision curve analysis through quantifying the net benefits at different threshold probabilities for the entire set [27]. The y-axis measured the net benefit, calculated by summing the benefits (true positive values) and subtracting the harms (false positive values). The x-axis represented the percentage of threshold probability, i.e., probability of positive clinical outcome (locoregional recurrence, distant metastasis, and death). "All" or "none" referred to the situations when all or none of the patients had positive clinical outcome (locoregional recurrence, distant metastasis, and death) [28].

Statistical analysis

Statistical analyses were conducted using SPSS (version 23.0, SPSS), MedCalc (version 15.0, MedClac), and statistical packages R (The R Founding: http://www.r-project.org; version 3.4.3). A t test or Mann–Whitney U test was used to compare numerical variables (age, erythrocyte count, neutrophil count, platelet count, serum levels of gynecological tumor markers, tumor size, rad-scores, DFS time) between different groups. The chi-square test was used to compare categorical variables (age categorized by 50, age categorized by 55, HPV, Figo stage, differentiation, lymph node metastases, number of lymph node metastasis, LVI, invasion depth, additional chemotherapy) between different groups. LASSO Cox regression was implemented using the "glmnet" and "rms" packages. The Kaplan–Meier survival analysis was performed using the "survival" package and the comparisons were done using the log log-rank test. The "Hmisc" package was used for comparisons between C-indexes. The "rmda" package was used for decision curve analysis. A two-sided p value less than 0.05 was indicative of statistically significant difference.

Results

Clinical characteristics

Detailed baseline characteristics of the patients are shown in Table 1. There were no significant differences in any of the clinical and pathological characteristics between the training set and the validation set (p range, 0.093 to 0.999). The median duration of the follow-up was 29 months (interquartile range, 16–40 months).

Radiomics feature selection and three radiomics scores' construction

In total, 1187 T2WI features, 971 DWI features, 1054 ADC features, and 1092 CET1WI features with ICC higher than 0.8 remained. Reproducibility of the extracted features is depicted in Fig S1. Owing to multi-scale tumor segmentation, three different radiomics scores (rad-score (VOI$^{\text{entire}}$), rad-score (VOI$^{+5\text{ mm}}$), and rad-score (VOI$^{-5\text{ mm}}$)) were respectively built through LASSO Cox regression analysis. Rad-score (VOI$^{\text{entire}}$) contained three T2WI features, one DWI feature, two ADC features, and three CET1WI features. Rad-score (VOI$^{+5\text{ mm}}$) contained one T2WI feature, two DWI features, one ADC feature, and four CET1WI features. Rad-score (VOI$^{-5\text{ mm}}$) contained two T2WI features and one ADC feature. The process of LASSO Cox regression is shown in Fig S2. The formulae of the three radiomics scores are detailed in Appendix 1. Detailed information regarding the final selected single radiomics features in each of the three radiomics scores is shown in Table S4.

DFS prediction performance of the three radiomics scores

The three radiomics scores showed favorable performances in DFS prediction with C-index values over

0.700, both in the training set and in the validation set. Rad-score (VOI$^{+5\text{ mm}}$) performed best, with a C-index of 0.750 (95% CI: 0.658, 0.842) in the training set and 0.839 (95% CI: 0.696, 0.983) in the validation set. Two presented ESSCC patients, who had distinctly different DFS times (49.6 months vs 13.8 months) with similar clinicopathological features, showed significantly different rad-score (VOI$^{\text{entire}}$) (-1.9 vs 1.3; $p < 0.001$), rad-score (VOI$^{+5\text{ mm}}$) (-1.7 vs 2.5; $p < 0.001$), and rad-score (VOI$^{-5\text{ mm}}$) (-2.3 vs 0.6; $p < 0.001$) (Fig. 3).

The performances of the final selected single radiomics features in the three radiomics scores for 2-year DFS probability are shown in Table S4. The distribution of three built rad-scores and their correlation with DFS are shown in Figure S3. The predictive performances of the three radiomics scores in predicting DFS and 2-year DFS probability are shown in Table S5.

Development, performance, and validation of clinical and combined model

Three clinicopathological characteristics (age categorized by 55, Figo stage, and LVI) and three established radiomics scores (rad-score (VOI$^{\text{entire}}$), rad-score (VOI$^{+5\text{ mm}}$), and rad-score (VOI$^{-5\text{ mm}}$)) were predictive of recurrence in univariable Cox analysis (Table 2). With application of stepwise multivariable Cox analysis with the lowest BIC score, age categorized by 55, Figo stage, LVI, and rad-score (VOI$^{+5\text{ mm}}$) were identified as independent predictors of recurrence (Table 2).
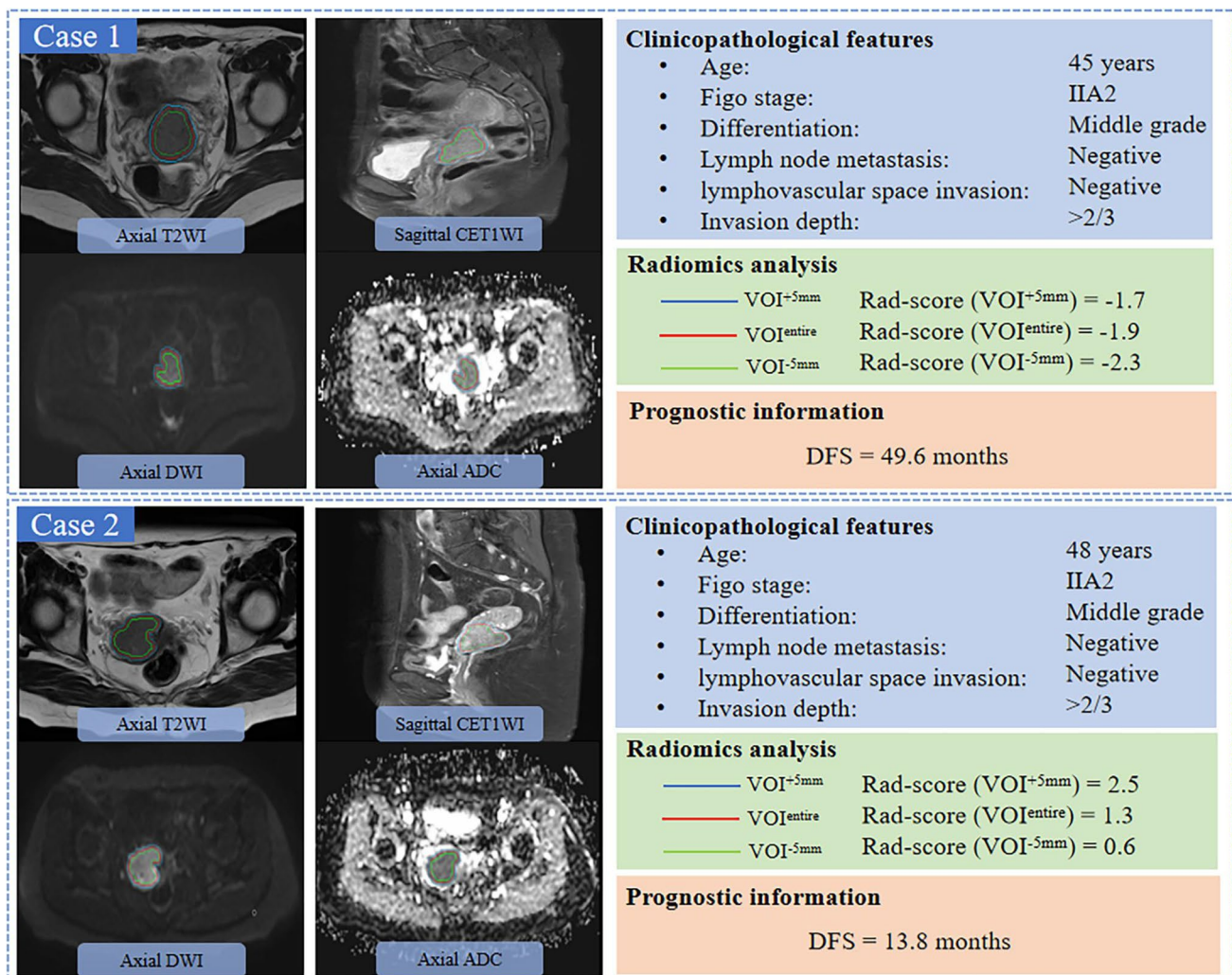

Fig. 3 Two presented cases of ESSCC patients who had distinctly different DFS time (49.6 months vs 13.8 months) with similar clinicopathological features showed significantly different rad-score (VOI$^{\text{entire}}$) (-1.9 vs 1.3; $p < 0.001$), rad-score (VOI$^{+5\text{ mm}}$) (-1.7 vs 2.5; $p < 0.001$), and rad-score (VOI$^{-5\text{ mm}}$) (-2.3 vs 0.6; $p < 0.001$)

Table 2 Univariable and multivariable Cox regression analyses for association with DFS

Variable	Univariable analysis		Multivariable analysis	
	Hazard ratio	*p* value	Hazard ratio	*p* value
Age categorized by 55	2.713 (1.393, 5.281)	0.003	1.251 (1.157, 1.352)	0.007
Figo stage	2.597 (1.275, 5.290)	0.009	1.744 (0.839, 3.623)	0.036
LVI	3.467 (1.225, 9.810)	0.019	4.106 (1.437, 11.732)	0.008
Rad-score (VOI^{entire})	1.129 (1.007, 1.352)	0.001	NA	NA
Rad-score (VOI$^{+5\ mm}$)	1.551 (1.357, 1.852)	< 0.001	1.278 (1.178, 1.386)	< 0.001
Rad-score (VOI$^{-5\ mm}$)	1.231 (1.057, 1.452)	< 0.001	NA	NA

Parentheses indicate 95% confidence interval

Figo, Federation of Gynecology and Obstetrics; *LVI*, lymphovascular space invasion; *NA*, not appliable

Fig. 4 The combined model was constructed and presented as a nomogram (**a**). To use the nomogram, add corresponding points for four variables (vertical projection to a point to the first graph scale for each variable) and locate it on the total points scale. Vertical project from the total points scale to the bottom risk scale and get the probability of DFS. The linear predictor scale is the coordinate value of the linear predictive value. Through corresponding transformation function, the linear predictive value can be converted to the probability of DFS. The graphs show areas under time-dependent ROC curves of the nomogram for DFS in the training set (**b**) and validation set (**c**). The calibration curves demonstrated good calibration of the nomogram in the training set (**d**) and validation set (**e**)
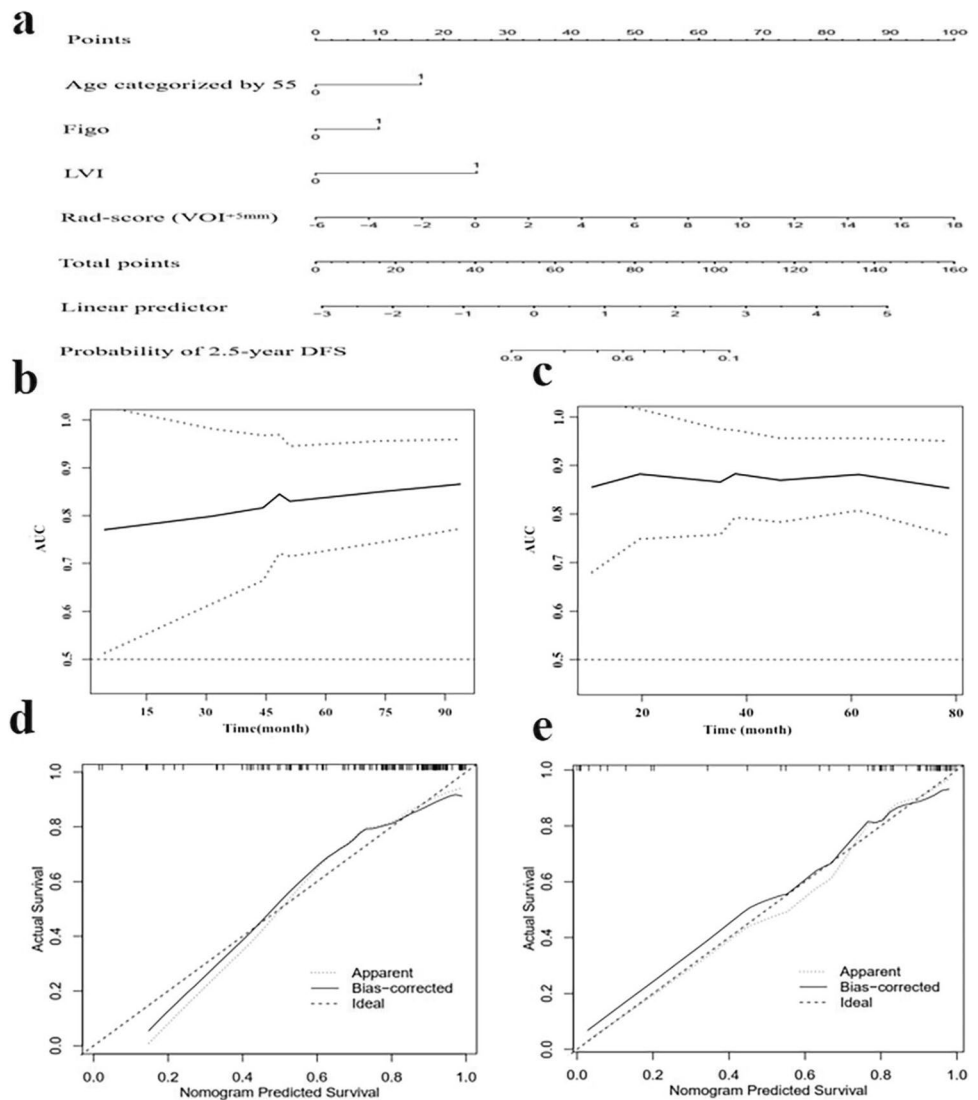
Age categorized by 55, Figo stage, and LVI were finally selected to build the clinical model. For DFS prediction, the clinical model possessed a *C*-index of 0.709 (95% CI: 0.624, 0.784) in the training set and 0.719 (95% CI: 0.583, 0.831) in the validation set. The combined model was constructed after combining these three clinicopathological characteristics with rad-score (VOI$^{+5\ mm}$) and was presented as a nomogram (Fig. 4a). The combined model demonstrated good predictive performance, with a *C*-index of 0.815 (95% CI: 0.739, 0.876) in the training set and 0.866 (95% CI: 0.748,

Table 3 Predictive performance of clinical and combined model in predicting DFS and 2-year DFS probability

Indicator	Clinical model		Combined model	
	Training set	Validation set	Training set	Validation set
C-index	0.709 (0.624, 0.784)	0.719 (0.583, 0.831)	0.815 (0.739, 0.876)	0.866 (0.748, 0.942)
AUC	0.725 (0.642, 0.798)	0.702 (0.565, 0.817)	0.843 (0.770, 0.900)	0.871 (0.754, 0.945)
Accuracy	0.641 (0.597, 0.761)	0.649 (0.577, 0.782)	0.793 (0.683, 0.886)	0.849 (0.721, 0.973)
Sensitivity	0.633 (0.543, 0.717)	0.718 (0.582, 0.877)	0.700 (0.606, 0.853)	0.827 (0.690, 0.940)
Specificity	0.700 (0.611, 0.772)	0.633 (0.579, 0.713)	0.876 (0.798, 0.932)	0.878 (0.782, 0.979)

Parentheses indicate 95% confidence interval

DFS, disease-free survival; *AUC*, area under curve

Table 4 Predictive performance of the combined model in predicting DFS at multiple time points

Time	AUC	Accuracy	Sensitivity	Specificity
1-year				
Training set	0.808 (0.732, 0.871)	0.771 (0.659, 0.895)	0.833 (0.516, 0.979)	0.764 (0.679, 0.836)
Validation set	0.853 (0.761, 0.892)	0.856 (0.737, 0.877)	0.875 (0.673, 0.997)	0.853 (0.757, 0.895)
1.5-year				
Training set	0.861 (0.791, 0.915)	0.789 (0.683, 0.903)	0.778 (0.577, 0.914)	0.833 (0.749, 0.898)
Validation set	0.853 (0.761, 0.892)	0.856 (0.737, 0.877)	0.875 (0.673, 0.997)	0.853 (0.757, 0.895)
2-year				
Training set	0.843 (0.770, 0.900)	0.793 (0.683, 0.886)	0.700 (0.606, 0.853)	0.876 (0.798, 0.932)
Validation set	0.871 (0.754, 0.945)	0.849 (0.721, 0.973)	0.827 (0.690, 0.940)	0.878 (0.782, 0.979)
2.5-year				
Training set	0.846 (0.774, 0.902)	0.807 (0.704, 0.907)	0.765 (0.588, 0.893)	0.822 (0.733, 0.913)
Validation set	0.862 (0.743, 0.939)	0.846 (0.639, 0.937)	0.867 (0.595, 0.983)	0.805 (0.651, 0.912)
3-year				
Training set	0.838 (0.765, 0.896)	0.811 (0.707, 0.889)	0.743 (0.567, 0.875)	0.820 (0.731, 0.890)
Validation set	0.866 (0.748, 0.942)	0.839 (0.653, 0.944)	0.875 (0.617, 0.984)	0.825 (0.672, 0.927)

Parentheses indicate 95% confidence interval

DFS, disease-free survival; *AUC*, area under curve

0.942) in the validation set. The combined model performed better in DFS prediction than the clinical model, both in the training set and in the validation set ($p = 0.024$; $p = 0.001$, respectively). The predictive performances of the clinical and the combined models for predicting DFS and 2-year DFS probabilities are summarized in Table 3 and shown in Fig S4. The predictive performances of the combined model in predicting DFS at multiple time points are detailed in Table 4. Time-dependent ROC curves of the combined model in DFS prediction are depicted in Fig. 4b–c. Good calibration of the combined model was confirmed by calibration curves (Fig. 4d–e).

Kaplan–Meier analysis

The Kaplan–Meier analysis was applied to further validate prognostic value of the selected clinicopathological and radiomics characteristics. Age > 55, Figo IIA, presence of LVI, and higher risk of rad-score (VOI$^{+5\,mm}$) were all significantly associated with worse DFS in the entire cohort ($p = 0.001$; $p = 0.001$; $p = 0.001$; $p < 0.001$, respectively). The Kaplan–Meier cumulative event curves for ESSCC recurrence stratified by age categorized by 55, Figo stage, LVI, and rad-score (VOI$^{+5\,mm}$) are depicted in Fig. 5.

Clinical practice

The combined model showed good positive net benefits across wide ranges of probability thresholds. When the threshold probability was 0.286 and 0.500, the net benefit value was 0.500 and 0.387, respectively. When the threshold probability was over 0.103, the combined model added more benefit than in the situations where all or none of the patients had positive clinical outcome and also more than the clinical model (Fig. 6). This demonstrated the favorable clinical utility of the combined model in aiding clinicians to determine adjuvant treatment (chemotherapy) in patients with ESSCC according to the risk of recurrence.
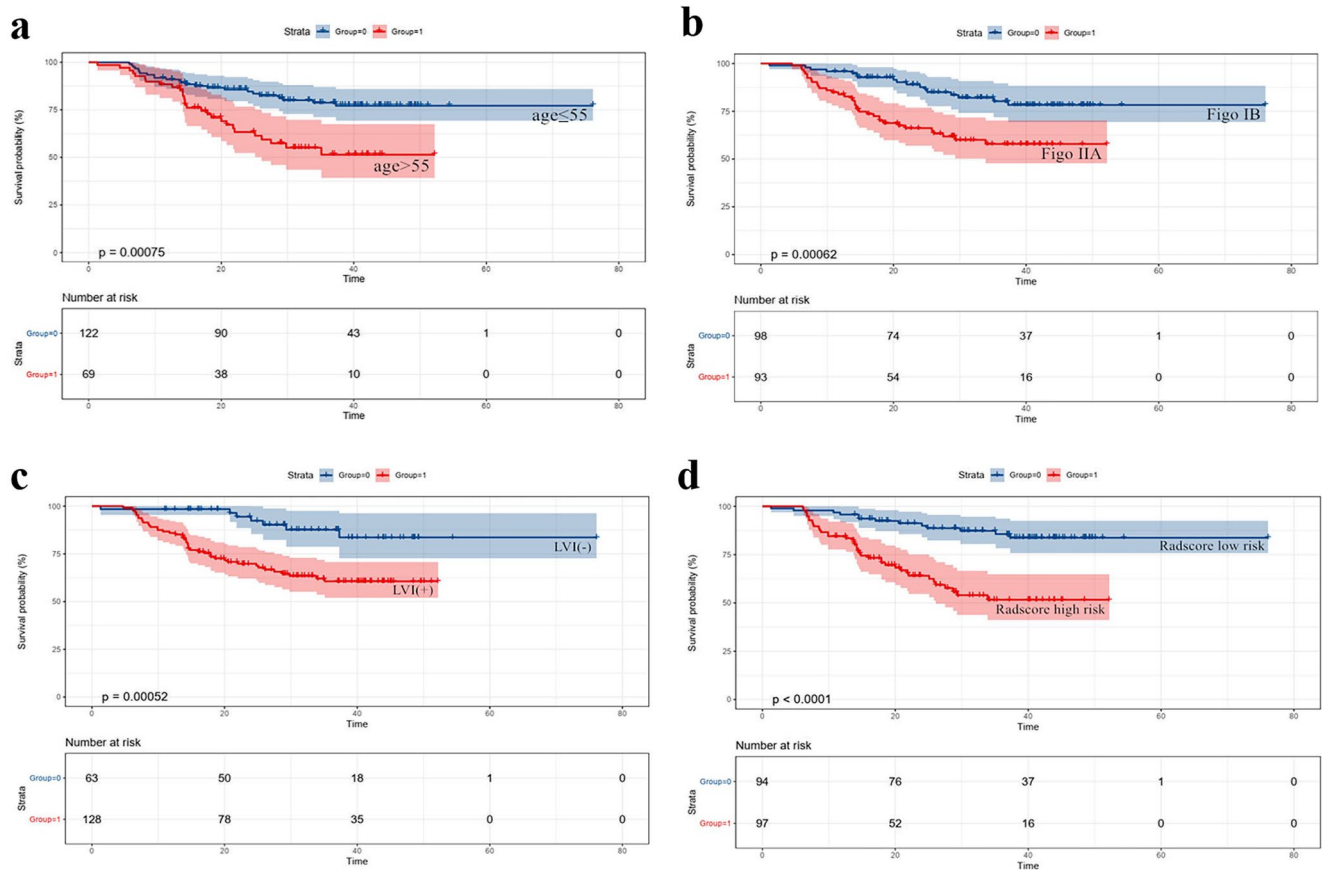
Fig. 5 The Kaplan–Meier cumulative event curves for survival in ESSCC stratified by age categorized by 55 (**a**), Figo stage (**b**), LVI (**c**), and rad-score (VOI$^{+5\,mm}$) (**d**)

Discussion

In this study, we conducted multiparametric MRI-derived radiomics based on multi-scale tumor region to predict DFS in ESSCC patients. Encouragingly, the combined model, which incorporated radiomics features with clinicopathological characteristics, exhibited improved prognostic performance and better clinical usefulness compared with the clinical model alone. The combined model presented as a nomogram can be easily used to predict survival, thereby facilitating the choice of initial surgery, postoperative adjuvant radiotherapy, and the intensity of postoperative observation.

Radiomics has been recognized as a useful imaging technology tool in predicting survival of patients with cervical cancer [16, 29]. Fang et al [16] verified an MRI-derived radiomics score as a prognostic biomarker in patients with early cervical cancer. Lucia et al confirmed PET/CT and MRI radiomics as independent predictors of recurrence in locally advanced cervical cancer treated with chemoradiotherapy [29]. However, previous studies have conducted tumor segmentation using a completely manual approach which was limited by poor inter-reader reliability. Besides that, feature extraction was based on the entire volumetric tumor region, while peritumoral information was ignored. Instead of the manual approach, we performed tumor segmentation semi-automatically using an integrated hybrid algorithm. This method was less affected by hand-related artifacts. Apart from the entire volumetric interest (VOI^{entire}) segmentation, the regions at a 5-mm distance outer (VOI$^{+5\,mm}$) and inner (VOI$^{-5\,mm}$) from the tumor surface were also reconstructed to obtain multi-scale tumor information. Consistent with other studies [16, 29], all of the three types of rad-scores all showed favorable performances in DFS prediction. These results indicate that radiomics may be a promising tool in assisting prediction of survival in ESSCC patients. Notably, rad-score (VOI$^{+5\,mm}$) performed best both in the training set and in the validation set. Moreover, rad-score (VOI$^{+5\,mm}$) was identified as an independent predictor of survival in multivariable Cox regression analysis. These results indicate that radiomics features from the peritumoral region reflecting the biological properties associated with intratumor heterogeneities [17, 18] may provide more information in prediction of cervical cancer prognosis.
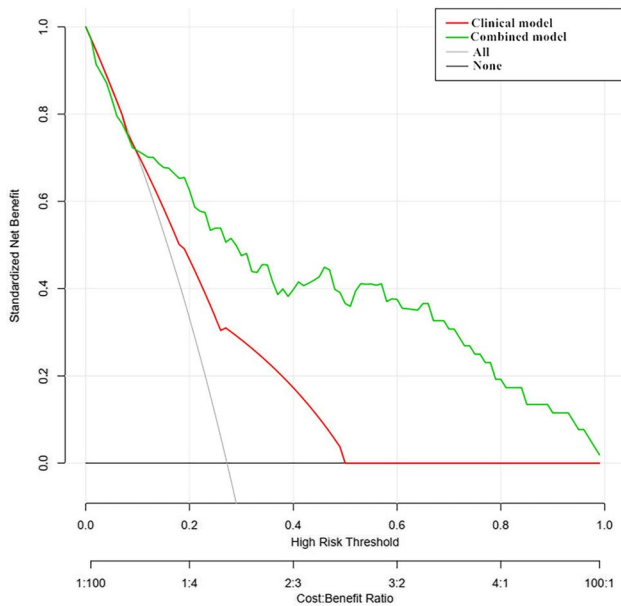
Fig. 6 Decision curve analysis for the clinical model and the combined model in predicting DFS. The *y*-axis stands for the net benefit, and the *x*-axis stands for the threshold probability. The decision curves indicated that if the threshold probability was over 0.103, the combined model (green line) added more benefit than the situation when all (gray line) or none (black line) of the patients had positive clinical outcome and more than the clinical model (red line)

Besides those from conventional structural images, radiomics features from DWI, ADC, and CET1WI were also analyzed in our study. In total, eight radiomics features (including one feature extracted from T2WI, two from DWI, one from ADC map, and four from CET1WI) were identified to be predictive of DFS in rad-score (VOI$^{+5 mm}$). Shape-based MajorAxisLength is a measurement of the largest axis length of the tumor enclosing ellipsoid [21]. A larger MajorAxisLength indicates a larger tumor size, which is consistent with previous studies that have shown that tumors with larger size have poor prognosis [5, 7]. Texture-based features (LargeAreaHighGrayLevelEmphasis, GrayLevelNonUniformity, DependenceNonUniformity, DependenceVariance, GrayLevelNonUniformity, LargeAreaHighGrayLevelEmphasis, and ZoneVariance) describe the distribution of voxel signal intensities, which correlates with tumor heterogeneity [21]. Our results are in agreement with previous studies that have reported that intratumor heterogeneity was tightly correlated with invasion and metastasis of ESSCC [12, 16, 29]. Importantly, a half of the selected features (4 of 8) were derived from CET1WI; they indicate that intratumoral and peritumoral tumor enhancing features that are potentially associated with tumor perfusion and vascularization may provide more prognostic information for ESSCC patients [30, 31].

By combining the clinical risk factors and radiomics features, we finally established the radiomics nomogram containing age categorized by 55, Figo stage, LVI, and rad-score (VOI$^{+5 mm}$) to predict DFS in ESSCC patients. Consistent with previous studies [4–6], older age, higher Figo stage, and existence of LVI were regarded as adverse prognostic factors. Notably, the radiomics nomogram possessed favorable predictive ability and it added significant incremental performance to the clinical model. Hence, it indicated the prognostic value of multiparametric MRI-derived radiomics in predicting DFS for ESSCC patients.

The current study had several limitations. First, this study was performed in a single center and the sample size was relatively small. Further multi-center studies with larger sample size are needed. Second, 32 of 191 patients received additional chemotherapy in addition to hysterectomy, which may have influenced survival. Third, some acquisition parameters differed from the current recommendations in Europe [7], which may cause potential bias. The slice thickness of T2WI and DWI was 5 mm, which was thicker than recommended (4 mm or less). Field of view of T2WI (240 × 240 mm) and DWI (380 × 380 mm) were not matched, which may have influenced side-by-side interpretation. Fourth, dynamic contrast-enhanced scan was not used; however, its derived parameters can potentially be used to identify tumor hypoxia regions tightly correlated with tumor migration, invasion, and metastasis [31]. This may have influenced the generalizability of our results. Fifth, in-depth tumor biological behavior was not analyzed and correlated with radiomics features. Radiomics feature explanation at a biological level should be conducted in further radiogenomics studies. Sixth, deep learning method was not applied due to the limited sample size. In the future, studies combining the use of deep learning and radiomics method for better prediction of DFS for ESSCC patients are needed.

In conclusion, multiparametric MRI-derived radiomics based on multi-scale tumor region is useful in predicting DFS in ESSCC patients, thereby facilitating clinical decision-making.

Declarations

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
● retrospective
● diagnostic or prognostic study
● performed at one institution

References

1. Cohen PA, Jhingran A, Oaknin A, Denny L (2019) Cervical cancer. Lancet 393:169–182
2. Banerjee S (2017) Bevacizumab in cervical cancer: a step forward for survival. Lancet 390:1626–1628
3. Vanichtantikul A, Tantbirojn P, Manchana T (2017) Parametrial involvement in women with low-risk, early-stage cervical cancer. Eur J Cancer Care (Engl) 26:1–5
4. Bhatla N, Berek JS, Cuello Fredes M et al (2019) Natarajan J. Revised FIGO staging for carcinoma of the cervix uteri. Int J Gynaecol Obstet 145:129–135
5. Biewenga P, van der Velden J, Mol BW et al (2011) Prognostic model for survival in patients with early stage cervical cancer. Cancer 117:768–776
6. Weyl A, Illac C, Lusque A et al (2020) Prognostic value of lymphovascular space invasion in early-stage cervical cancer. Int J Gynecol Cancer 30:1493–1499
7. Manganaro L, Lakhman Y, Bharwani N et al (2021) Staging, recurrence and follow-up of uterine cervical cancer using MRI: updated guidelines of the European Society of Urogenital Radiology after revised FIGO staging 2018. Eur Radiol 31(10):7802–7816
8. Xiao M, Yan B, Li Y, Lu J, Qiang J (2020) Diagnostic performance of MR imaging in evaluating prognostic factors in patients with cervical cancer: a meta-analysis. Eur Radiol 30:1405–1418
9. Wang YT, Li YC, Yin LL, Pu H (2016) Can diffusion-weighted magnetic resonance imaging predict survival in patients with cervical cancer? A meta-analysis. Eur J Radiol 85:2174–2181
10. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A (2015) Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res 21:249–257
11. Savadjiev P, Chong J, Dohan A et al (2019) Image-based biomarkers for solid tumor quantification. Eur Radiol 29:5431–5440
12. Umutlu L, Nensa F, Demircioglu A et al (2020) Radiomics analysis of multiparametric PET/MRI for N- and M-staging in patients with primary cervical cancer. Rofo 192:754–763
13. Wang T, Gao T, Guo H et al (2020) Preoperative prediction of parametrial invasion in early-stage cervical cancer with MRI-based radiomics nomogram. Eur Radiol 30:3585–3593
14. Li Z, Li H, Wang S et al (2019) MR-based radiomics nomogram of cervical cancer in prediction of the lymph-vascular space invasion preoperatively. J Magn Reson Imaging 49:1420–1426
15. Xiao M, Ma F, Li Y et al (2020) Multiparametric MRI-based radiomics nomogram for predicting lymph node metastasis in early-stage cervical cancer. J Magn Reson Imaging 52:885–896
16. Fang J, Zhang B, Wang S et al (2020) Association of MRI-derived radiomic biomarker with disease-free survival in patients with early-stage cervical cancer. Theranostics 10:2284–2292
17. Hu Y, Xie C, Yang H et al (2020) 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 3:e2015927
18. Xu X, Zhang HL, Liu QP et al (2019) Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol 70:1133–1144
19. Wels MG, Lades F, Muehlberg A, Suehling M (2019) General purpose radiomics for multi-modal clinical research. Proc SPIE 1095046
20. Moltz JH, Bornemann L, Kuhnigk JM et al (2009) Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans. IEEE J Sel Top Sign Proces 3:122–134
21. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77:e104–e107
22. Zhang Z (2016) Variable selection with stepwise and best subset approaches. Ann Transl Med 4:136
23. O'brien RM, (2007) A caution regarding rules of thumb for variance inflation factors. Qual Quant 41:673–690
24. Bland JM, Altman DG (1998) Survival probabilities (the Kaplan-Meier method). BMJ 317:1572
25. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B (2011) survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics 27:3206–3208
26. Heagerty PJ, Lumley T, Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 56:337–344
27. Fitzgerald M, Saville BR, Lewis RJ (2015) Decision curve analysis. JAMA 313:409–410
28. Xie G, Wang R, Shang L et al (2020) Calculating the overall survival probability in patients with cervical cancer: a nomogram and decision curve analysis-based study. BMC Cancer 20:833
29. Lucia F, Visvikis D, Desseroit MC et al (2018) Prediction of outcome using pretreatment 18F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. Eur J Nucl Med Mol Imaging 45:768–786
30. Willmott LJ, Monk BJ (2009) Cervical cancer therapy: current, future and anti-angiogensis targeted treatment. Expert Rev Anticancer Ther 9:895–903
31. Hauge A, Wegner CS, Gaustad JV, Simonsen TG, Andersen LMK, Rofstad EK (2017) DCE-MRI of patient-derived xenograft models of uterine cervix carcinoma: associations with parameters of the tumor microenvironment. J Transl Med 15:225