European Radiology (2019) 29:3358–3371
https://doi.org/10.1007/s00330-019-06124-9

IMAGING INFORMATICS AND ARTIFICIAL INTELLIGENCE

Magnetic resonance imaging radiomics in categorizing ovarian masses and predicting clinical outcome: a preliminary study

He Zhang[1] · Yunfei Mao[2] · Xiaojun Chen[3] · Guoqing Wu[2] · Xuefen Liu[1] · Peng Zhang[1] · Yu Bai[4] · Pengcong Lu[5] · Weigen Yao[5] · Yuanyuan Wang[2] · Jinhua Yu[2,6] · Guofu Zhang[1]

Received: 21 September 2018 / Revised: 9 February 2019 / Accepted: 22 February 2019 / Published online: 8 April 2019
© European Society of Radiology 2019

Abstract

Purpose To evaluate the ability of MRI radiomics to categorize ovarian masses and to determine the association between MRI radiomics and survival among ovarian epithelial cancer (OEC) patients.

Method A total of 286 patients with pathologically proven adnexal tumor were retrospectively included in this study. We evaluated diagnostic performance of the signatures derived from MRI radiomics in differentiating (1) between benign adnexal tumors and malignancies and (2) between type I and type II OEC. The least absolute shrinkage and selection operator method was used for radiomics feature selection. Risk scores were calculated from the Lasso model and were used for survival analysis.

Result For the classification between benign and malignant masses, the MRI radiomics model achieved a high accuracy of 0.90 in the leave-one-out (LOO) cross-validation cohort and an accuracy of 0.87 in the independent validation cohort. For the classification between type I and type II subtypes, our method made a satisfactory classification in the LOO cross-validation cohort (accuracy = 0.93) and in the independent validation cohort (accuracy = 0.84). Low-high-high short-run high gray-level emphasis and low-low-high variance from coronal T2-weighted imaging (T2WI) and eccentricity from axial T1-weighted imaging (T1WI) images had the best performance in two classification tasks. The patients with higher risk scores were more likely to have poor prognosis (hazard ratio = 4.1694, $p = 0.001$).

Conclusion Our results suggest radiomics features extracted from MRI are highly correlated with OEC classification and prognosis of patients. MRI radiomics can provide survival estimations with high accuracy.

Key Points

- *The MRI radiomics model could achieve a higher accuracy in discriminating benign ovarian diseases from malignancies.*
- *Low-high-high short-run high gray-level emphasis, low-low-high variance from coronal T2WI, and eccentricity from axial T1WI had the best performance outcomes in various classification tasks.*
- *The ovarian cancer patients with high-risk scores had poor prognosis.*

Keywords Ovarian epithelial cancer · Magnetic resonance imaging · Computer-assisted diagnosis · Radiomics

He Zhang and Yunfei Mao contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06124-9) contains supplementary material, which is available to authorized users.

✉ Jinhua Yu
jhyu@fudan.edu.cn

✉ Guofu Zhang
guofuzh@fudan.edu.cn

[1] Department of Radiology, Obstetrics and Gynecology Hospital, Fudan University, Shanghai, People's Republic of China

[2] Department of Electronic Engineering, Fudan University, Shanghai, People's Republic of China

[3] Department of Gynecology, Obstetrics and Gynecology Hospital, Fudan University, Shanghai, People's Republic of China

[4] Center for Child and Family Policy, Duke University, Durham, NC, USA

[5] Department of Radiology, Yuyao People's Hospital, Ningbo, Zhejiang province, People's Republic of China

[6] Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention, Shanghai, People's Republic of China

Abbreviations

ADC	Apparent diffusion coefficient
DWI	Diffusion-weighted magnetic resonance imaging
FIGO	International Federation of Gynecology and Obstetrics
ISR	Iterative sparse representation
OEC	Ovarian epithelial cancer
PACS	Picture archiving and communication system
SRC	System sparse representation coefficient
SVM	Support vector machine

Introduction

Ovarian cancer is the most malignant gynecological disease, accounting for 22,500 deaths annually in China [1]. Ovarian epithelial cancer (OEC) is the most frequently occurring gynecological malignancy, representing more than 70% of deaths in patients with ovarian disease [2, 3]. Pathologically, OEC is approximately divided into two subtypes: type I and type II ovarian cancers as each has different treatments and prognoses [4, 5]. The significance of distinguishing two OEC subtypes is that type I has different biological characteristics, treatment response, and gene mutations from type II [3]. Type II ovarian cancers (most are high-grade serous ovarian cancers) often couple with chromosomal instability and have aggressive biological behavior with poorer prognoses [4]. To accurately classify patients into the correct category, noninvasive imaging methods may be able to guide clinicians to design specialized treatment plans.

With superb soft tissue resolution and no radiation, magnetic resonance imaging (MRI) is widely used as a problem-solving modality in clinical units to suggest the etiology of adnexal lesions, which cannot be determined by ultrasound or computed tomography (CT) [6, 7]. With multiparametric MRI imaging methods (for example, diffusion-weighted imaging and dynamic-enhanced MRI), MRI can provide a quantitative index to further differentiate malignant conditions from benign ovarian tumors [8–11]. Despite these advantages, it is still difficult to identify OEC subtypes with MRI alone.

Although there have been a few studies focused on discriminating type I and type II subtypes, preoperative diagnosis is challenging due to some overlaps in both clinical indexes and imaging characteristics [12, 13]. In a large cohort study, patients with type I tumors were confirmed to have an increased overall survival after 2 years of follow-up compared with those with type II tumors [14]. Considering the pathological subtypes, patients with clear cell tumors presented a worse survival rate than those with other subtypes do [15, 16]. Apart from the International Federation of Gynecology and Obstetrics (FIGO) stage as an independent predictor, optimal cytoreduction is also considered as an important predictor of overall survival for ovarian cancer patients [16].

Radiomics is defined as a new "data-driven" approach for extracting large sets of innumerable quantitative features from tomographic images (CT, MR, or positron emission tomography images) [17]. With sophisticated image processing methods, all clinical medical images are transferred to mineable high-throughput image features, which thereafter can be used to correlate these processed feature signatures with pathology diagnoses or treatment responses [18]. Using a multimodal approach, recent studies have shown that quantitative parameters derived from multiparametric MRI radiomics could be used as imaging biomarkers to assess tumor presence, spread, recurrence, and response to treatment in female cancer patients [19]. To date, there have been limited MRI radiomics studies concerning ovarian cancer, especially for ovarian cancer subtype categorization. Therefore, improving the capability of diagnosing ovarian cancer subtypes and identifying MRI radiomics signatures for prognosis estimation were the motives for this study. The purpose of this research was threefold: first, we planned to evaluate the diagnostic performance of the MRI radiomics model in discriminating benign ovarian tumors from malignancies; second, we sought to establish a model to validate whether MRI radiomics could differentiate between type I and type II OEC and then to assess the correlation between MRI radiomics results and histological findings; third, we conducted a survival analysis to determine whether MRI radiomics features could determine patient prognosis.

Patients and methods

Patients

Our institutional review board (Gynecological and Obstetric Hospital, School of Medicine, Fudan University, Shanghai, China) approved this retrospective study, and the requirement for informed consent was waived for all participants. From January 2014 to December 2017, 438 consecutive patients with clinically suspected gynecological diseases were retrospectively retrieved from our Institutional Picture Archiving and Communication System (PACS, GE). The inclusion criteria were as follows: (1) no previous pelvic surgery; (2) no previous gynecological disease history; and (3) MRI examinations before pelvic or laparoscopic surgery were performed at our institution. The exclusion criteria were as follows: (1) previous pelvic surgical history or radiation history; (2) MRI data were unavailable either for the examination performed at another institution or due to claustrophobia; and (3) no histological results. Finally, the total number of patients who met the criteria was 280 (average age, 52.7 ± 12.3 years). The sample comprised 100 patients with type I cancer, 81 patients with type II cancer, 27 patients with other ovarian malignancies, and 72 patients with benign etiologies

(Table 1 and Fig. 1). All included patients were pathologically proven laparoscopy or laparotomy. The information on FIGO stage, pathological type, immunohistological staining results, and laboratory tests were collected through a hospital information system.

Patient follow-up

All patients were followed up every 6 months during the first 3 years, and annually thereafter. We used disease-free survival as the endpoint. The time range was defined as the number of days between the first day of treatment to the date of disease progression (either determined by imaging or clinical examinations), death, or the date of the last follow-up inquiry. Patient follow-up was mainly performed by telephone inquiry (P.Z.). All information was provided by patients themselves or their relatives who knew their medical history.

MR acquisition and interpretation

MRI was performed using a 1.5-T MR system (Magnetom Avanto, Siemens) with a phased-array coil. The routine MRI protocols used for the assessment of pelvic masses included axial turbo spin-echo (TSE) T1-weighted imaging (T1WI), sagittal TSE T2-weighted imaging (T2WI), and axial/sagittal TSE fat-suppressed T2WI (fs-T2WI). The detailed MRI acquisition parameters are listed in Supplementary Table 1. On $b = 800$ mm^{-2}/s diffusion-weighted imaging (DWI) images, apparent diffusion coefficient (ADC) values were measured manually on a post-processing workstation (Leonardo, Siemens) by one operator (X.F.L.). Regions of interest were drawn centrally in both cystic and solid areas with no more than three sites in each lesion on $b = 800$ mm^{-2}/s DWI images. Round or elliptical shapes were centrally placed in the targeted region with an area range of 160–220 mm^2. Only the lowest ADC value was used for subsequent statistical analysis. For any lesions that were not clearly depicted on DWI images, axial fs-T2WI images and dynamic contrast-enhanced (CE) images were also reviewed to help to define the small nodular components. CE pelvic imaging was acquired at multiple phases of contrast medium enhancement in both sagittal and axial planes. The contrast medium was manually injected into the elbow vein.

Two observers (G. Z. and H. Z, with 20 and 15 years of experience in gynecological imaging, respectively), who were blinded to the histological results, independently analyzed each participant's MRI dataset in the testing group on a PACS terminal server. The etiology of adnexal masses on MRI was determined based on previously well-described definitions in the literature [11, 20, 21]. For interobserver discrepancies in both the detection and characterization of adnexal lesions, consensus was achieved or one decision (G. Z.) was accepted based on more experience. Confidence in diagnosing lesion presence was provided using a five-point scale as follows: "5" means that the lesion is definitely malignant; "4," the lesion is probably malignant; "3," uncertain; "2," the lesion is probably benign; and "1," the lesion is definitely benign. Cases with one or more lesions that were identified as probably or definitely malignant were considered positive; all other confidence ratings were considered negative.

Table 1 The summary of the pathological types and numbers of the selected samples

Pathological type	Numbers	Age (years)*
Type I ovarian cancer	98	43.2 ± 14.0
LGSC	16	
EC	5	
Clear cell carcinoma	26	
Borderline serous/mucinous cystadenomas	51	
Type II ovarian cancer		55.0 ± 10.5
HGSC	85	
Others malignancies	28	42.9 ± 15.3
Granulosa cell tumor	7	
Dysgerminoma	2	
Metastatic ovarian cancer	10	
Yolk sac tumor	2	
Immature teratoma	4	
Malignant mixed germ cell tumor	3	
Benign etiologies	75	44.5 ± 17.3
Serous/mucinous cystadenoma	62	
Fibroma/fibrothecoma	6	
Mature teratoma	2	
Endometrioma	1	
Struma ovarii	1	
Total	286	

*Mean ± standard deviation; *LGSC*, low-grade serous carcinoma; *EC*, endometrioid carcinoma; *HGSC*, high-grade serous carcinoma

MRI lesion segmentation and radiomics features selection

All lesion segmentation was performed by an experienced radiologist (H.Z.). We chose one slice with the largest lesion diameter in each protocol as the premium picture for segmenting the whole lesion. The segmentation was manually outlined using MATLAB software (version R2015b, MathWorks, Natick) (Fig. 2). After the tumor segmentation process, radiomics features were extracted from four modal pictures. For axial T1WI, sagittal fs-T2WI, and coronal T2WI MRI images, the obtained features mainly fell into four groups: intensity, shape, texture, and wavelets. The intensity group consisted of 22 features that described the overall intensity and heterogeneity information in the segmented area. The shape group consisted of 15 features, capturing the external boundary structure of the lesion. The texture group contained 39 features, estimating the gray-level regional

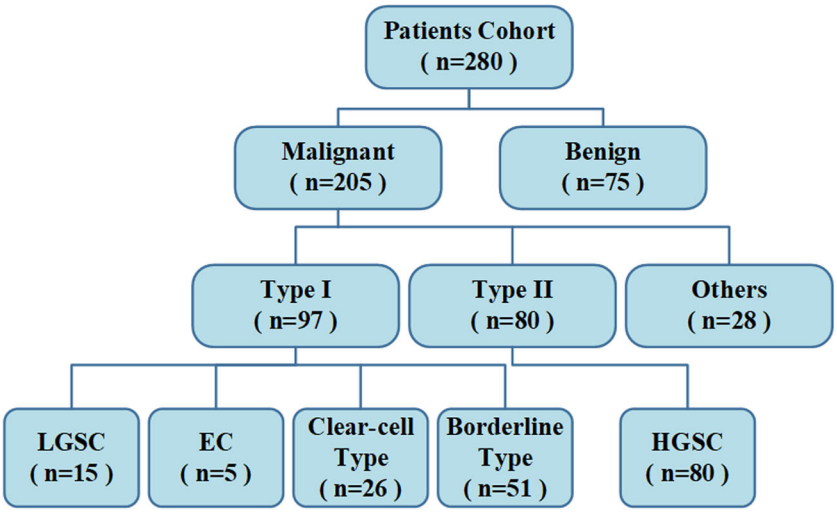


Fig. 1 Flow chart of patient selection for different classification tasks

spatial distribution. Compared with the intensity features, the texture features focused more on the locality in addition to frequency [22, 23]. In the wavelet group, we transformed the intensity and texture features into eight frequency subbands via wavelets to obtain more additional unseen information. The calculation details for these four sets of features have been given in previous research [24]. In total, we extracted 1714 features from four types of MRI acquisition images for each lesion. The details of the high-throughput features are summarized in Supplementary Table 2.

Classification task analysis

To obtain the most effective feature subset, we applied the iterative sparse representation (ISR)–based feature selection method for all extracting features [25, 26]. The sparse representation coefficient (SRC) denoting the importance of the corresponding feature is calculated on average. Support vector machine (SVM) is a widely applied classifier and is efficient in machine-learning tasks with limited samples [27]. The details of the ISR feature selection and SVM are given in Supplementary Appendix A. The whole dataset was randomly divided into two parts: a cross-validation cohort and an independent validation cohort (Table 2). The leave-one-out (LOO) cross-validation was first applied to test the MRI radiomics classification model. Then, the independent validation set was used to further evaluate the diagnostic performance of this model.

Survival analysis

By observing the results in Kaplan-Meier (KM) plot, we obtained the MRI radiomics signatures most related to the patients' survival status. In the multivariate analysis, the least absolute shrinkage and selection operator (Lasso) method is widely used as a selection method for high-dimensional data [28–30]. The Lasso regression is applied in the prediction

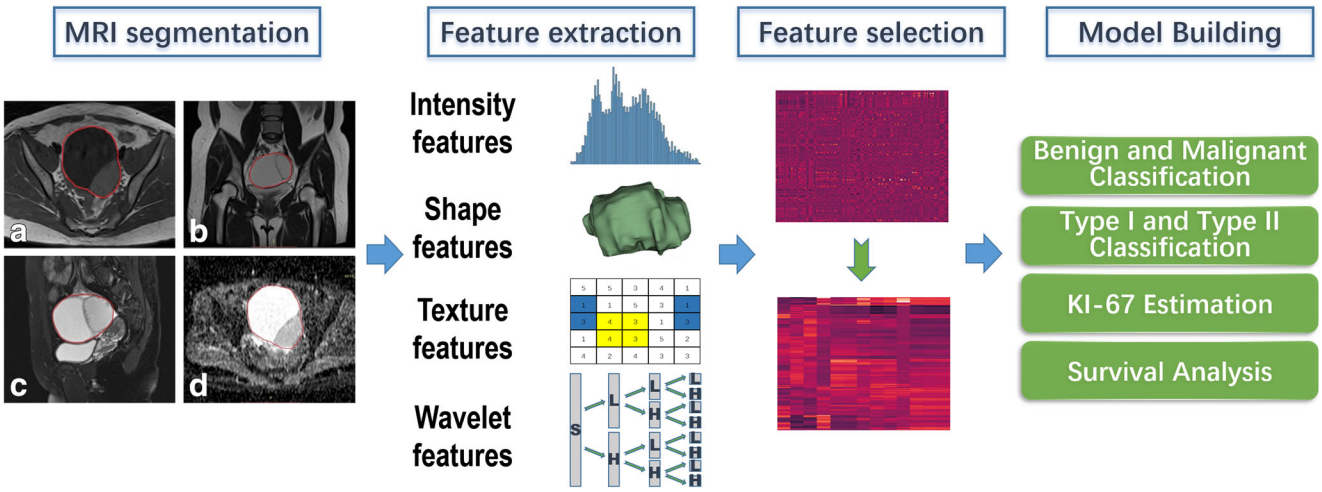


Fig. 2 The MRI radiomics analysis process from extraction to model building. MRI imaging segmentation sample in four sequences for a single patient. A) Axial T1 WI, B) Coronal T2WI, C) Sagittal fs-T2WI, D) ADC map

Table 2 Clinical and pathological data summaries in both cross-validation and independent validation cohort

	LOO cross-validating cohorts (numbers)[§] ($N = 195$)	Independent validation cohorts (numbers) ($N = 85$)	p value
Age (years)	47.08 ± 15.81	44.66 ± 16.26	0.884
< 30	35 (17.9%)	30 (35.3%)	
30–50	66 (33.8%)	18 (21.1%)	
> 50	94 (48.3%)	37 (43.6%)	
Ki-67 expression (%)	28.28 ± 23.20	22.72 ± 21.91	0.073
< 50	139 (71.3%)	70 (82.4%)	
50–75	46 (23.6%)	8 (9.4%)	
> 75	10 (5.1%)	7 (8.2%)	
CA-125 level (IU/L)	483.69 ± 821.00	513.04 ± 1105.27	0.555
< 35	28 (14.4%)	25 (29.4%)	
35–200	72 (36.9%)	32 (37.6%)	
200–500	35 (17.9%)	15 (17.6%)	
> 500	60 (30.8%)	13 (15.4%)	
Category			0.113
Type 1	60 (30.8%)	37 (47.1%)	
LGSC	9 (4.7%)	6 (7.1%)	
EC	1 (0.5%)	4 (4.7%)	
Clear Cell type	17 (8.7%)	9 (11.8%)	
Borderline type	33 (16.9%)	18 (23.5%)	
Type 2			
HGSC	63 (32.3%)	17 (21.2%)	
Benign	52 (26.7%)	23 (23.5%)	
Malignant	20 (10.3%)	8 (8.2%)	
FIGO*			0.111
I	44 (35.7%)	26 (44.8%)	
II	14 (11.4%)	2 (3.4%)	
III	55 (44.7%)	27 (46.5%)	
IV	10 (8.1%)	3 (5.2%)	

[§] Indicates one LGSC lesion and five HGSC lesions cannot be outlined during the segmentation; *FIGO stage was recorded in Type I and II cancer group. *LOO*, leave-one-out

operation, where the output is defined as the risk score. The calibration plot was generated in order to evaluate the performance of nomogram. The data processing work is based on the MATLAB R2012, SPSS Statistics 25 (IBM Corporation) and R software (R Foundation for Statistical Computing).

Results

Comparing the diagnostic performance of discriminating malignancies from benign tumors between radiologist and computer

The computerized program performed better than the radiologists did in distinguishing malignant adnexal masses from benign tumors. MRI radiomics using all of the sequences yielded an accuracy (ACC) of 90.6%, which is higher than that of the radiologists with 83.5%. The most common diagnostic error among the radiologists was classifying borderline ovarian tumors (BOTs) into the benign group. Regarding BOTs as low potential malignancies, these tumors often received conservative therapy, which is different from the regimens used for OEC. When BOTs were excluded, the radiologist's discriminating ability competed with the computer's performance (Table 3); however, MRI radiomics still showed the best performance. In this regard, interestingly, all diagnostic mistakes by radiologists were omitted by computers and vice versa.

ADC values were not available in three cases in the benign group either due to small tumor size or severe pulsation artifacts due to large size. Overall, the ADC values showed significant differences between benign

Table 3 Diagnostic performance comparison between radiologist and computer in determining malignant adnexal lesions in validation group at reference standard in the patient-based evaluation (95% confidence interval (CI))

	N	TP	TN	FP	FN	SEN	SPE	PPV	NPV	ACC
Reader's diagnostic performance (include BOT)	85	51	20	3	11	0.823 (0.700–0.904)	0.869 (0.653–0.966)	0.944 (0.837–0.986)	0.645 (0.454–0.802)	0.835 (0.736–0.904)
Computer with all selected sequence	85	56	21	2	6	0.903 (0.795–0.960)	0.913 (0.705–0.985)	0.966 (0.871–0.994)	0.778 (0.573–0.906)	0.906 (0.816–0.956)
Reader's diagnostic performance (not include BOT)	67	42	20	3	2	0.955 (0.833–0.992)	0.869 (0.653–0.966)	0.933 (0.807–0.983)	0.909 (0.694–0.984)	0.925 (0.828–0.972)
Computer with all selected sequence	67	42	21	2	2	0.955 (0.833–0.992)	0.913 (0.705–0.985)	0.955 (0.833–0.992)	0.913 (0.705–0.985)	0.940 (0.847–0.981)

N = TP + TN + FP + FN

TP, true positive; *TN*, true negative; *FP*, false positive; *FN*, false negative; *AUC*, area under the ROC curve; *PPV*, positive predictive values; *NPV*, negative predictive values

and malignant tumors and between type I and type II cancers. If the ADC value was set to 1162×10^{-3} mm^2/s, the sensitivity (SEN) was 85.1%, the specificity (SPE) was 83.0%, and the area under the receiver operating characteristic (ROC) curve (AUC) was 91.4% for differentiating benign from malignant tumors. For discriminating type I from type II cancer, with a cut-off ADC value of 924×10^{-3} mm^2/s, the SEN, the SPE, and AUC were 72.7%, 62.5%, and 73.7%, respectively.

Classification results based on MRI radiomics features

Table 4 illustrates the final classification results by the LOO cross-validation and the independent validation groups. The model was first determined on the LOO cross-validation experiment based on the AUC. Then, we evaluated the model on the independent validation set. Our results yielded an ACC of 87% and an AUC of 95% of in differentiating between benign and malignant ovarian tumors. For discriminating type I and type II ovarian cancer subtypes, our model achieved an ACC of 83% and an AUC of 85%. When we used the median value of Ki-67 expression as a cut-off point, an AUC of 73% and an accuracy of 75% were achieved in the independent validation group. The ROC curves for three classification tasks in the independent validation group are summarized in Fig. 3.

To evaluate the individual performance of each sequence, we applied the same classification method based on a single sequence. When we used the features extracted from a single modality, the diagnostic performance was inferior to that extracted from the multi-modality models. The results and the number of features contributed by each sequence are shown in Table 5. Overall, T1WI performed best in benign versus malignancy classification. Moreover, sagittal fs-T2WI was the most suitable sequence for type I versus type II classification. Considering the three classification tasks, the contributed numbers from three modalities (T1WI, coronal T2WI, and sagittal fs-T2WI) were similar: the ADC map contributed more features to categorizing ovarian benign and malignant ovarian tumors than the two other classification tasks did.

The most important radiomics features were low-high-high (LHH), short-run high gray-level emphasis (SRHGE) from coronal T2WI for identifying benign and malignant tumors ($p < 0.000$), low-low-high (LLH) variance from coronal T2WI images for classifying type I and type II ($p < 0.000$) (Supplementary Table 3), and eccentricity from axial T1WI images for determining Ki-67 expression 50% threshold classification ($p < 0.000$) (Fig. 4).

Table 4 The diagnostic performance of all selected MRI radiomics features for various classification tasks

	Classification	AUC	ACC	SEN	SPE
LOO cross-validation cohort	Benign and malignancy	0.9746	0.9026	0.9441	0.7885
	Type I and type II	0.8593	0.9268	0.9048	0.9500
	KI-67 (threshold 50)	0.8057	0.7879	0.8830	0.5526
Independent testing cohort	Benign and malignancy	0.9670	0.9059	0.9032	0.9130
	Type I and type II	0.8228	0.8333	0.7647	0.8649
	KI-67 (threshold 50)	0.7377	0.7541	0.8400	0.5000

LOO, leave-one-out; *AUC*, area under the receiver operator curve; *ACC*, accuracy; *SENS*, sensitivity; *SPEC*, specificity

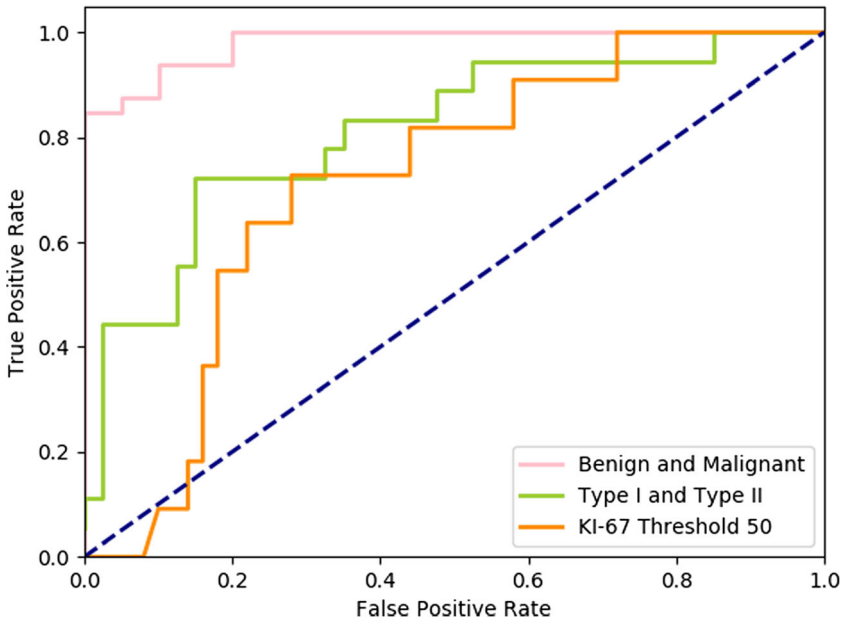


Fig. 3 Receiver operating characteristic (ROC) curves of the three classification tasks based on MRI radiomics

Radiomics features related to patients' follow-up

The best useful MRI radiomics features selected for survival analysis are listed according to their *p* values (Supplementary Table 4). When the mean was used as a cut-off value, the patients with lower LHL small zone emphasis (LHL-SZE) values had much poorer survival than those with high LHL-SZE values (34% versus 7.5%) did. T1WI radiomics features were more likely associated with the clinical outcome than the other protocols were. The survival analyses based on MRI radiomics features and clinical characteristics are shown in KM plots (Fig. 5). After the features were scored by the Lasso method (Supplementary Appendix C), our model divided patients into good and poor prognosis groups (Fig. 5) with an AUC of 0.899. Patients with higher-risk scores were more likely to suffer from disease progression or death after treatment (HR = 4.362, $p = 0.0013$) (Fig. 6). The concordance index (*C*-index) for the nomogram that integrated the radiomics signature with the clinical characteristics was 0.755. This *C*-index value was higher than the value of 0.603 from the nomogram with clinical data alone or 0.683 from the nomogram with only radiomics features (Supplementary Table 5). The best features with low *p* value in the single radiomics feature prediction analysis were not selected into the multiple features set (Supplementary Appendix B).

Discussion

By using conventional MRI characteristics and quantitative parameters generated from functional MRI, some previous

Table 5 Diagnostic performance of MRI radiomics in each modality in three classification tasks and the summaries of contribution feature numbers for each modality

Task		Modalities				
		T1WI	Coronal T2WI	Sagittal fs–T2WI	ADC map	Total features
Benign and malignant	AUC	0.9243	0.9123	0.8705	0.7756	
	ACC	0.8588	0.8260	0.8117	0.7765	
	Feature contribution	23	30	26	5	84
Type I and type II	AUC	0.7647	0.7678	0.8292	0.7407	
	ACC	0.7222	0.7222	0.7857	0.7027	
	Feature contribution	18	16	21	1	56
KI-67 (threshold 50)	AUC	0.7273	0.7273	0.6337	0.6109	
	ACC	0.7037	0.6439	0.6066	0.5574	
	Feature contribution	7	5	3	0	15

AUC, area under the receiver operator curve; *ACC*, accuracy

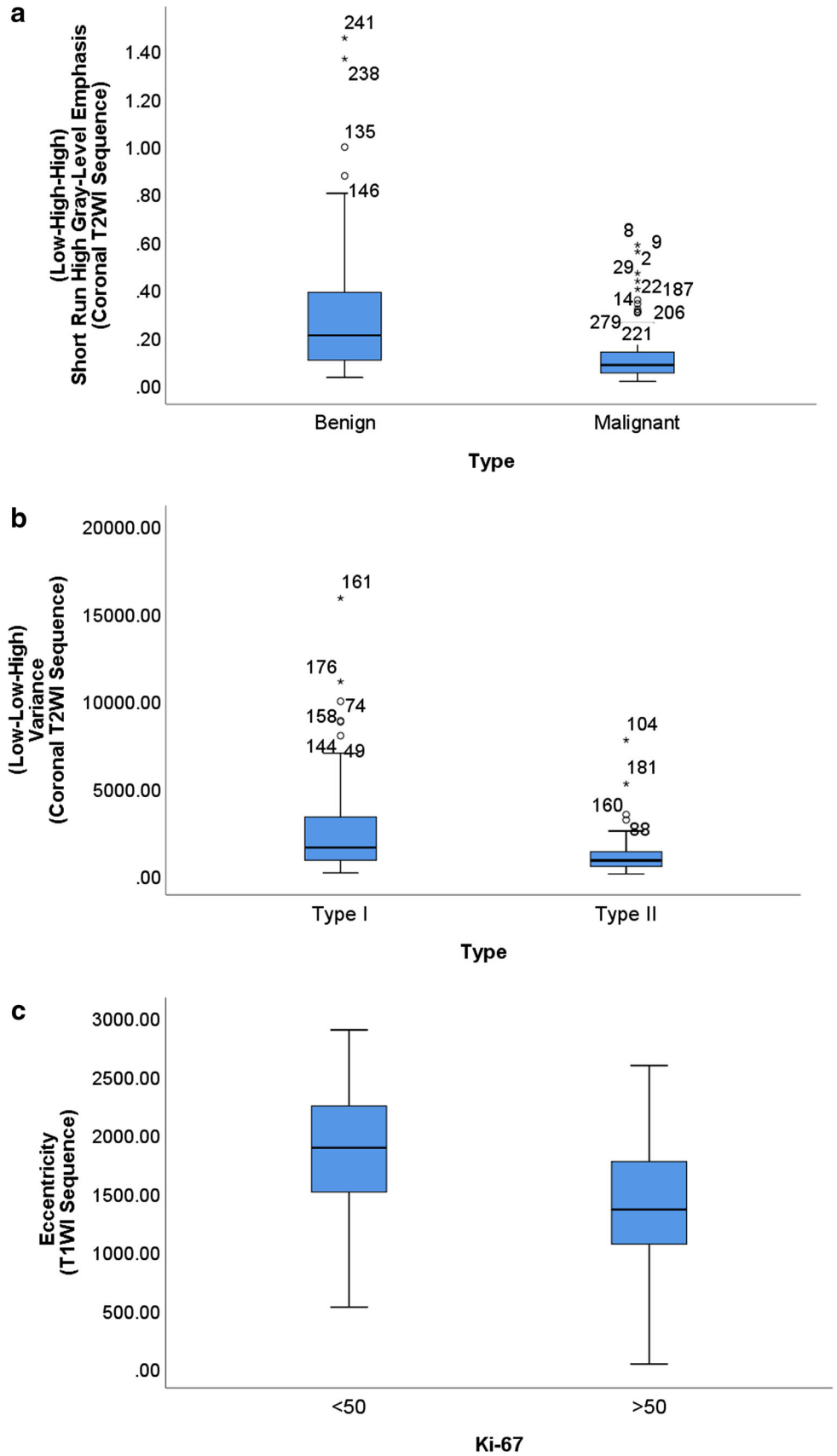
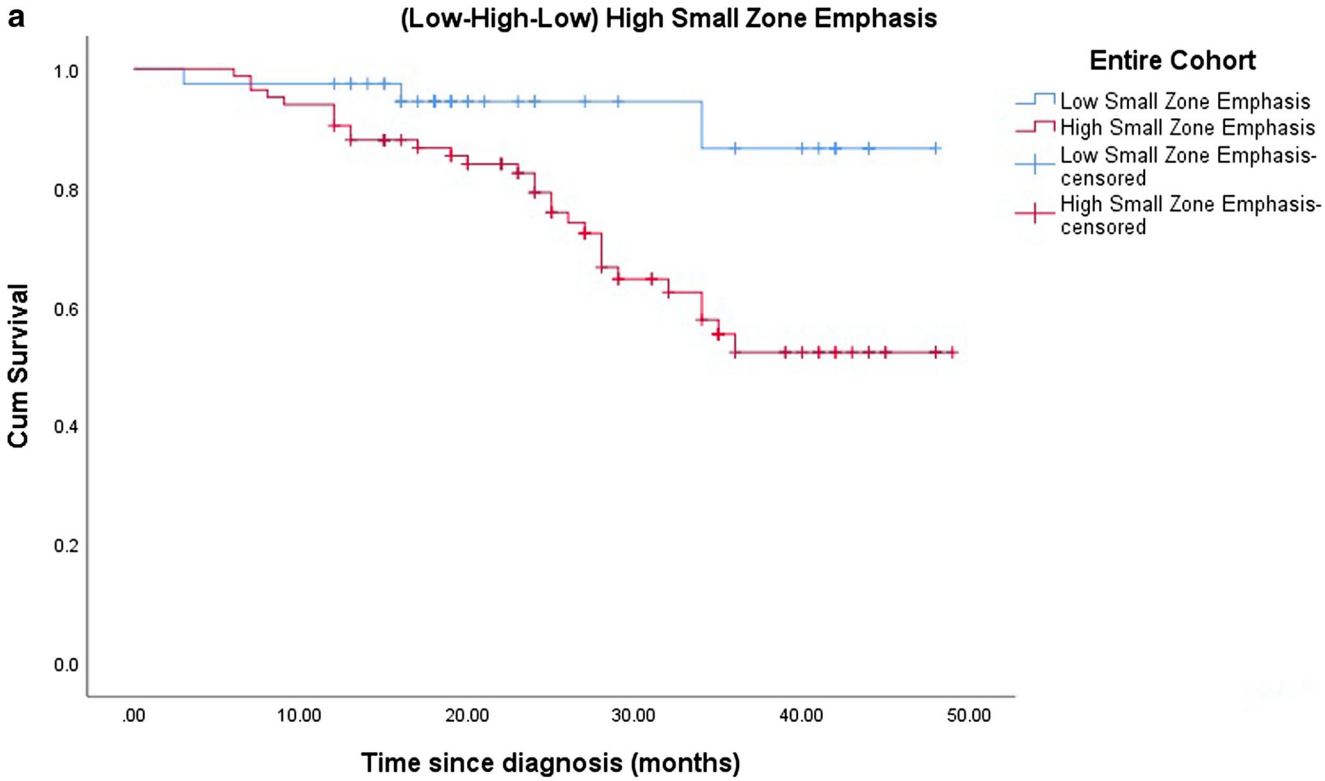


Fig. 4 Boxplots of the best MRI radiomics features in each classification task. **a** Benign and malignant classification. **b** Type I and type II classification. **c** 50% threshold Ki-67 classification



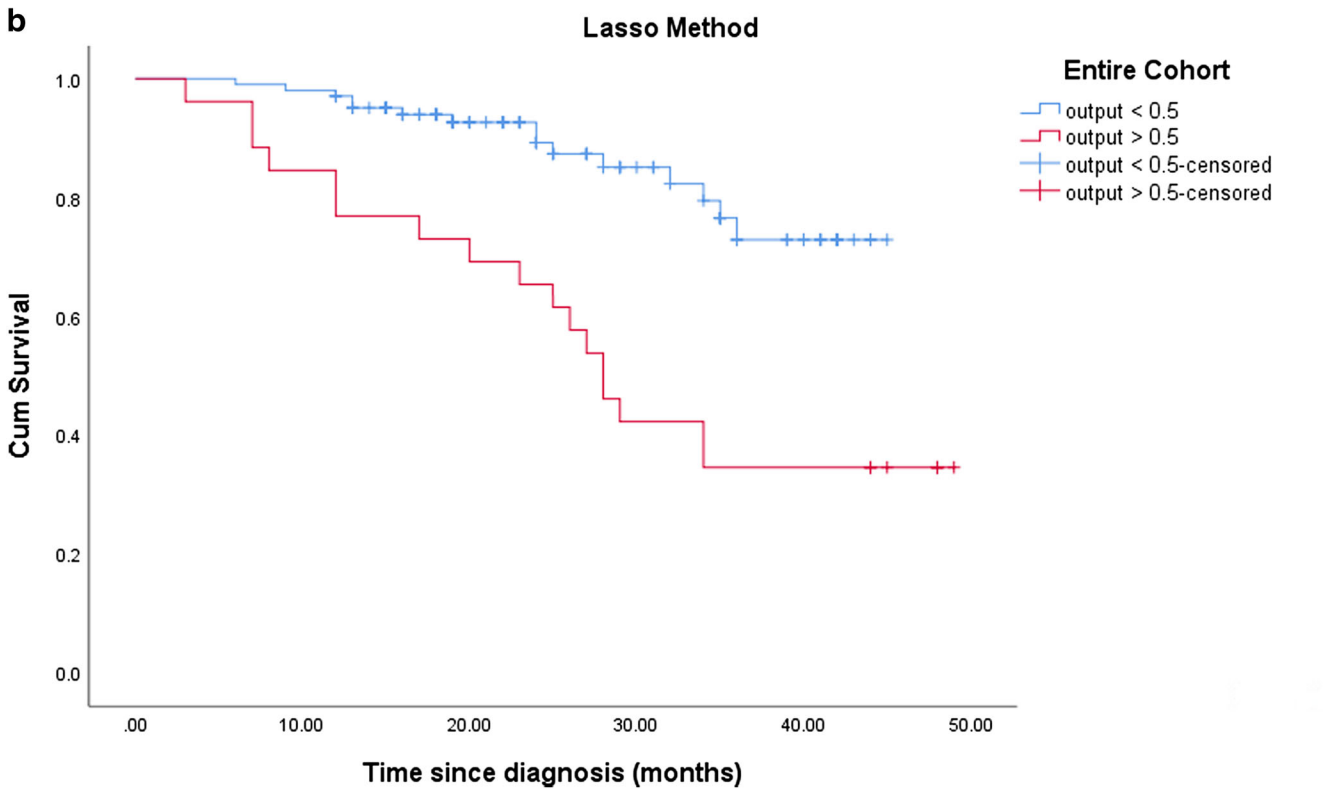


Fig. 5 Survival analysis with various MRI radiomics features. **a** Low-high-low small zone emphasis feature. **b** Lasso method. **c** Ovarian epithelial cancer type. **d** Age

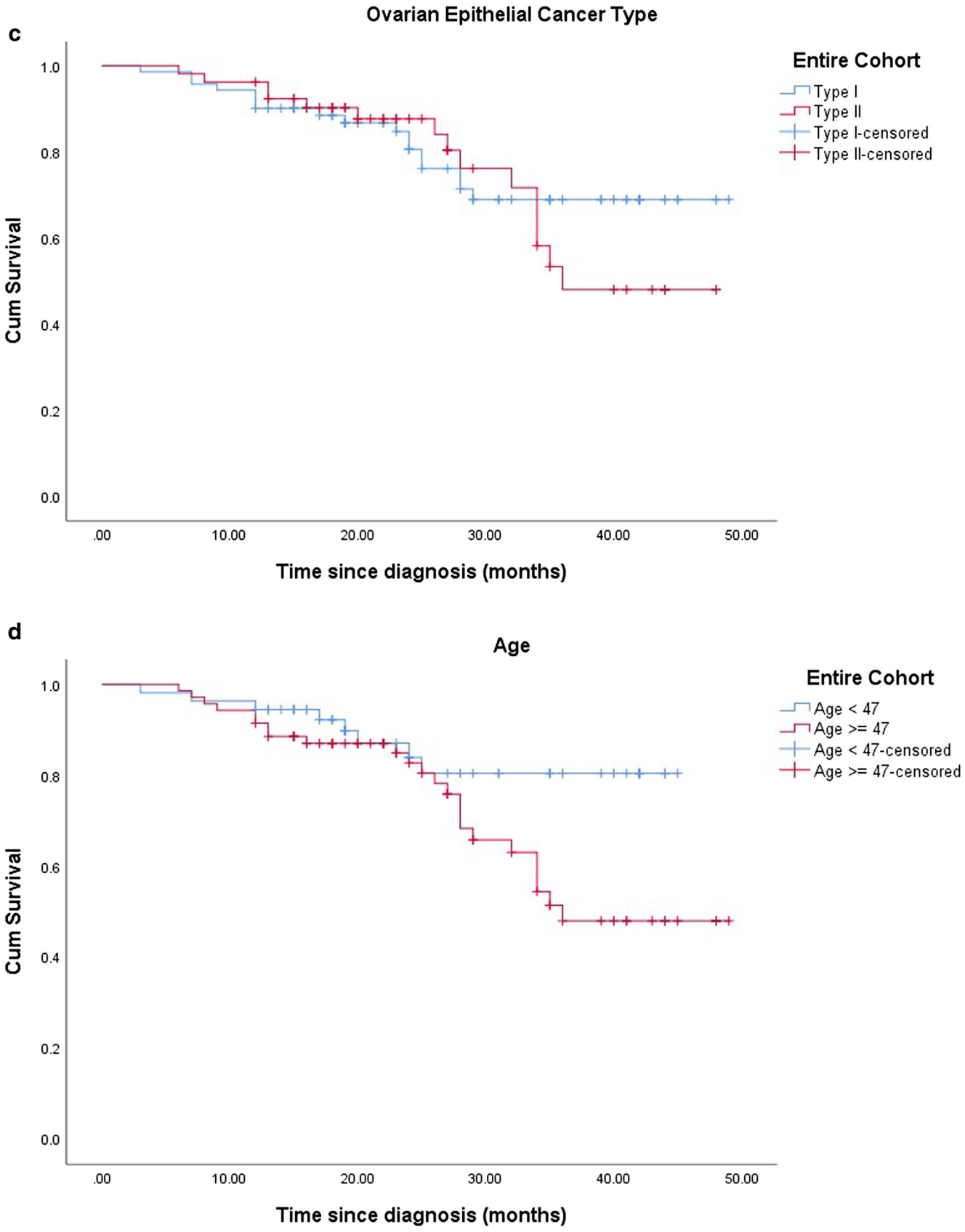


Fig. 5 (continued)

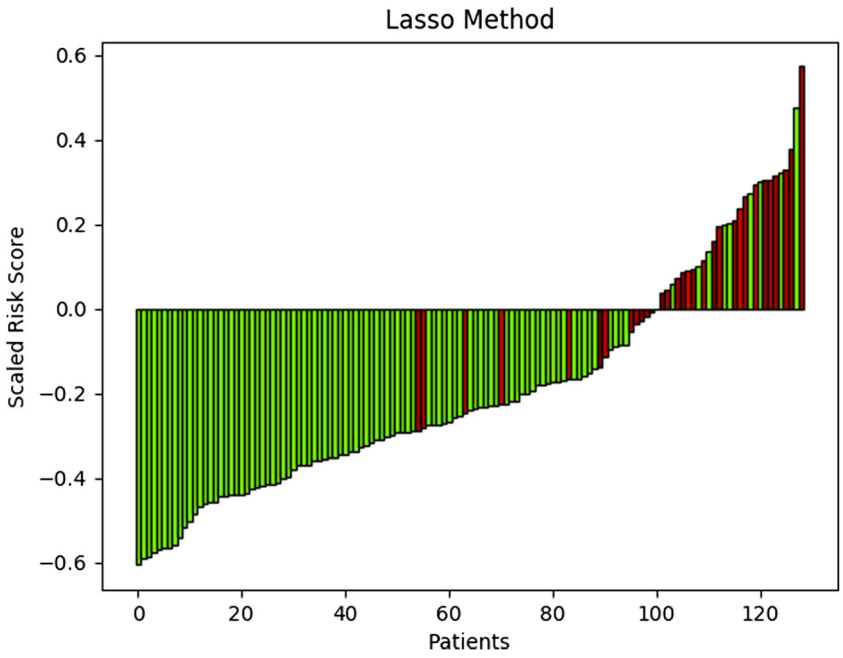


Fig. 6 Risk scores for each patient and their true survival status. The red bars represent the patients who suffered disease progression or died (the dark red represents the patients who experienced the disease progression in a short time). The green bars belong to the patients without any recorded events during follow-up

studies have shown promising results in both ovarian tumor classification and monitoring treatment response [5, 9, 27, 31, 32]. Our results demonstrated that this computed model yielded an excellent performance in differentiating benign and malignant ovarian tumors (AUC = 97.5% and ACC = 90.3%) and in classifying type I and type II ovarian cancers (AUC = 85.9% and ACC = 92.7%) in the cross-validation cohort. These findings were also confirmed in the independent cohort. In addition, the MRI radiomics signatures demonstrated a strong ability to stratify patients into a disease-free group and a disease-progression group with little overlaps. To the best of our knowledge, this is the first study using a large cohort with MRI radiomics in preoperative ovarian tumor diagnosis and treatment follow-up.

Several studies have used computer-aided diagnosis for ovarian cancer classification. Fathi et al reported their results in 55 sonographically indeterminate ovarian masses by using descriptive parameters in dynamic with linear discriminant analysis (LDA) and SVM classifiers [27]. They found that the time-to-peak and wash-in-rate parameters showed a high SEN (89% for LDA and 97% for SVM) and a high SPE (93% for LDA and 100% for SVM) in distinguishing malignancies from benign ovarian conditions [27]. Rizzo et al studied the potential associations between CT radiomics features and prognostic factors in a sample of 101 patients with high-grade serous carcinoma. Their results showed that F2-Shape/Max3Ddiameter radiomics feature below the median predicted a high risk of relapse in ovarian cancer patients during 1-year post-treatment (odds ratio = 11.86) [33]. Qiu et al acquired two sets of CT images (pre-treatment and 4–6-week post-treatment) to compare three image features (tumor volume, density, and density variance) between the two image sets in 30 ovarian cancer patients (12 responders and 18 non-responders) and to predict 6-month progress-free survival (PFS). Their model achieved an AUC of 0.831 in predicting PFS when combining all three features together [34].

Although some algorithms have been proposed for computer-aided diagnosis of ovarian cancer, there are no machine-learning models reported for type I and type II OEC classification. In turn, it is difficult for radiologists to make decisions by using MRI images alone. In our studied samples, MRI radiomics showed better performance than radiologists in discriminating ovarian cancer from benign tumors. Considering that the cases of diagnostic error between computer and radiologist are different, MRI radiomics may help radiologists to improve the diagnostic ACC before invasive procedures.

In contrast to previous studies that used a small number of image features [27, 33, 34], we calculated 1714 features from four MRI protocols for each lesion. This approach provides a better discriminative ability for tumor classification tasks [35]. In the present study, our model produced an AUC of 85% and an ACC of 84% in discriminating type I and type II OEC in the validation group. This finding has not yet been reported in previous studies yet. Recent studies performing MRI radiomics in gynecological tumor diagnosis and histological grade evaluation mainly relied on one set of MRI modalities [36, 37]. In this study, we selected four types of MRI modalities to establish an MRI radiomics model. The differences among the individual performance of fs-T1, sagittal fs-T2WI, and coronal T2 radiomics features in the classification task are insignificant.

Our results demonstrated that radiomics features from four jointed sequences yielded a better diagnostic performance in the

classification tasks than those from a single sequence did. Because the four MR acquisition sequences and the four related radiomics groups were included in our study, the number of total extracted features was large. We used an ISR method to obtain a set of the most important characteristics for ovarian cancer classification. As a useful tool for feature reduction, the ISR model returns each feature's SRC value, which is able to weight the importance of corresponding features. Therefore, we can use it to determine the optimal feature set for various classification tasks. The portion of the contributed features in each modality did not vary significantly among the three different classification tasks. T1WI radiomics contributed more useful features [18] for categorizing type I and type II cancers than T2WI did [16]. The possible reason could be that more signal differences were observed on T1WI images (high signal on T1WI) than on T2WI images. Furthermore, numerous studies have reported that ADC measurements could be used as additional tools for differentiating between benign and malignant conditions and could be applied as imaging biomarkers for monitoring treatment response [10, 12, 38, 39]. In the present study, the ADC value also differentiated benign from malignant tumors, which had much lower measurements. Because only intensity information in the ADC map is analyzed in clinical practice, we extracted 22 intensity features for analysis from ADC map. In the recent work, histogram features in the ADC map showed effectiveness in distinguishing between low-grade and high-grade cervical cancer [40]. Compared with the conventional MRI sequence, only a small number of features in the ADC map were captured in our study (564 versus 22, respectively). We found that the best feature for predicting tumor malignancy was the SRHGE feature from coronal T2WI images, which is similar to the long-run high gray-level emphasis (LRHGE) feature used in a previous radiomics analysis [41].

For the survival analysis, we used the Lasso method to determine the radiomics signatures from the total feature set. This approach was reported in a previous study that used radiomics features as prognostic factors in advanced nasopharyngeal carcinoma [41]. The Lasso model is a suitable method for analyzing a small sample with high-dimensional features due to its advantage of avoiding overfitting [42]. Our multivariate radiomics model successfully classified patients into high-risk and low-risk groups by using the risk score threshold. MRI radiomics signatures provided more accurate prognosis predictions than other clinical information did such as OEC type and patient age. Some previous studies showed the potential application of radiomics in survival analysis [33, 34]. In our model, the prediction performance based on multivariate radiomics feature analysis is mostly superior to that based on univariate analysis mostly. The explanation could be that the selected features from the Lasso method combined all related MRI radiomics signatures as a whole into consideration, while it was not considered during the univariate analysis process [43].

Several limitations should be noted. First, we did not include CE-MRI images to establish the MRI radiomics model. Although many studies have reported that CE-MRI is helpful for differentiating benign tumors from malignancies, the CE-MRI was not available for all included patients in the current study. Therefore, we did not select this protocol for analysis in order to diminish the selection bias. Second, lesion segmentation was manually outlined on a single slice (two-dimensional space) by one radiologist, and its evaluation may be biased toward different operators and volume segmentation. Finally, all MRI images were acquired in a 1.5-T MRI scanner; however, 3.0-T MRI machines are widely used in tertiary clinical centers. Further study is needed to explore the difference between the two modalities.

In summary, our results suggest that radiomics features extracted from MRI were highly correlated with ovarian cancer classification and patient prognosis. MRI radiomics features may provide ovarian cancer diagnosis and prognosis estimation with high ACC.

Funding This work is financially supported by the Shanghai Emerging Advanced Technology Joint Research Project (SHDC12014130).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Guofu Zhang.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry Statistician Yu Bai kindly provided all statistical work for this study.

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

Ethical approval Institutional Review Board approval was obtained.

Methodology
- Retrospective
- Observational
- Performed at one institution

References

1. Chen WZR, Baade PD, Zhang S et al (2016) Cancer statistics in China, 2015. CA Cancer J Clin 66:115–132

2. Cannistra SA (2004) Cancer of the ovary. N Engl J Med 351:2519–2529
3. Bowtell DD, Böhm S, Ahmed AA et al (2015) Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. Nat Rev Cancer 15:668–679
4. Lalwani N, Prasad SR, Vikram R, Shanbhogue AK, Huettner PC, Fasih N (2011) Histologic, molecular, and cytogenetic features of ovarian cancers: implications for diagnosis and treatment. Radiographics 31:625–646
5. Menon U, Griffin M, Gentry-Maharaj A (2014) Ovarian cancer screening—current status, future directions. Gynecol Oncol 132:490–495
6. Medeiros LR, Freitas LB, Rosa DD et al (2011) Accuracy of magnetic resonance imaging in ovarian tumor: a systematic quantitative review. Am J Obstet Gynecol 204:67–e1
7. Sala E, Wakely S, Senior E, Lomas D (2007) MRI of malignant neoplasms of the uterine corpus and cervix. AJR Am J Roentgenol 188:1577–1587
8. Li W, Chu C, Cui Y, Zhang P, Zhu M (2012) Diffusion-weighted MRI: a useful technique to discriminate benign versus malignant ovarian surface epithelial tumors with solid and cystic components. Abdom Imaging 37:897–903
9. Mitchell CL, O'Connor JPB, Jackson A et al (2010) Identification of early predictive imaging biomarkers and their relationship to serological angiogenic markers in patients with ovarian cancer with residual disease following cytotoxic therapy. Ann Oncol 21:1982–1989
10. Zhang H, Zhang G-F, He Z-Y, Li Z-Y, Zhang G-X (2014) Prospective evaluation of 3T MRI findings for primary adnexal lesions and comparison with the final histological diagnosis. Arch Gynecol Obstet 289:357–364
11. Thomassin-Naggara I, Balvay D, Aubert E et al (2012) Quantitative dynamic contrast-enhanced MR imaging analysis of complex adnexal masses: a preliminary study. Eur Radiol 22:738–745
12. Stein EB, Wasnik AP, Sciallis AP, Kamaya A, Maturen KE (2017) MR imaging–pathologic correlation in ovarian cancer. Magn Reson Imaging Clin N Am 25:545–562
13. Alcázar JL, Utrilla-Layna J, Mínguez JÁ, Jurado M (2013) Clinical and ultrasound features of type I and type II epithelial ovarian cancer. Int J Gynecol Cancer 23:680–684
14. Prahm KP, Karlsen MA, Høgdall E et al (2015) The prognostic value of dividing epithelial ovarian cancer into type I and type II tumors based on pathologic characteristics. Gynecol Oncol 136:205–211
15. Tang H, Liu Y, Wang X et al (2018) Clear cell carcinoma of the ovary: clinicopathologic features and outcomes in a Chinese cohort. Medicine (Baltimore) 97:e10881
16. Lee JY, Kim S, Kim YT et al (2018) Changes in ovarian cancer survival during the 20 years before the era of targeted therapy. BMC Cancer 18:601
17. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
18. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749
19. Anastasi E, Gigli S, Ballesio L, Angeloni A, Manganaro L (2018) The complementary role of imaging and tumor biomarkers in gynecological cancers: an update of the literature. Asian Pac J Cancer Prev 19:309–317
20. Rajkotia K, Veeramani M, Macura KJ (2006) Magnetic resonance imaging of adnexal masses. Top Magn Reson Imaging 17:379–397
21. Thomassin-Naggara I, Daraï E, Cuenod CA, Rouzier R, Callard P, Bazot M (2008) Dynamic contrast-enhanced magnetic resonance imaging: a useful tool for characterizing ovarian epithelial tumors. J Magn Reson Imaging 28:111–120
22. Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D (2018) Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. Eur Radiol 28:4514–4523
23. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
24. Yu J, Shi Z, Lian Y et al (2017) Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. Eur Radiol 27:3509–3522
25. Wu G, Chen Y, Wang Y et al (2018) Sparse representation-based radiomics for the diagnosis of brain tumors. IEEE Trans Med Imaging 37:893–905
26. Li H, Zhu Y, Burnside ES et al (2016) Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. NPJ Breast Cancer 2:16012
27. Kazerooni AF, Malek M, Haghighatkhah H et al (2017) Semiquantitative dynamic contrast-enhanced MRI for accurate classification of complex adnexal masses. J Magn Reson Imaging. https://doi.org/10.1002/jmri.25359
28. Meng Y, Zhang Y, Dong D et al (2018) Novel radiomic signature as a prognostic biomarker for locally advanced rectal cancer. J Magn Reson Imaging 48:605–614
29. Zhu X, Dong D, Chen Z et al (2018) Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. Eur Radiol 28:2772–2778
30. Ferreira Junior JR, Koenigkam-Santos M, Cipriano FEG, Fabro AT, Azevedo-Marques PM (2018) Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. Comput Methods Programs Biomed 159:23–30
31. Lindgren A, Anttila M, Rautiainen S et al (2017) Primary and metastatic ovarian cancer: characterization by 3.0T diffusion-weighted MRI. Eur Radiol 27:4002–4012
32. Wang F, Wang Y, Zhou Y (2017) Comparison between types I and II epithelial ovarian cancer using histogram analysis of monoexponential, biexponential, and stretched-exponential diffusion models. J Magn Reson Imaging 46:1797–1809
33. Rizzo S, Botta F, Raimondi S et al (2018) Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months. Eur Radiol 28:4849–4859
34. Qiu Y, Tan M, McMeekin S et al (2016) Early prediction of clinical benefit of treating ovarian cancer using quantitative CT image feature analysis. Acta Radiol 57:1149–1155
35. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ (2015) Machine learning methods for quantitative radiomic biomarkers. Sci Rep 5:13087
36. Liu Y, Zhang Y, Cheng R et al (2019) Radiomics analysis of apparent diffusion coefficient in cervical cancer: a preliminary study on histological grade evaluation. J Magn Reson Imaging 49:280–290
37. Ueno Y, Forghani B, Forghani R et al (2017) Endometrial carcinoma: MR imaging–based texture model for preoperative risk stratification—a preliminary analysis. Radiology 284:748–757
38. Marth C, Reimer D, Zeimet AG (2017) Front-line therapy of advanced epithelial ovarian cancer: standard treatment. Ann Oncol 28:viii36–viii39
39. Nakamura K, Joja I, Nagasaka T et al (2012) The mean apparent diffusion coefficient value (ADCmean) on primary cervical cancer

is a predictive marker for disease recurrence. Gynecol Oncol 127: 478–483
40. Schob S, Meyer HJ, Pazaitis N et al (2017) ADC histogram analysis of cervical cancer aids detecting lymphatic metastases—a preliminary study. Mol Imaging Biol 19:953–962
41. Zhang B, Tian J, Dong D et al (2017) Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 23:4259–4269
42. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S (2016) Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. BMC Med Res Methodol 16:154
43. Huang Y, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. Jpn J Clin Oncol 34:2157–2164

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