BMC Medical Imaging

RESEARCH

Open Access

An integrated radiomics and deep learning model on multisequence MRI for preoperative prediction of lymphovascular space invasion in endometrial cancer

Hong-Jian Luo[1†], Jin Cheng[2†], Ke Wang[2], Jia-Liang Ren[3], Li Mei Guo[4], Jinliang Niu[4] and Xiao-Li Song[4*]

Abstract

Purpose To develop and validate a multimodal model that integrates radiomics features (RFs) and deep learning features (DFs) derived from preoperative multisequence magnetic resonance imaging (MRI) for the prediction of lymphovascular space invasion (LVSI) in patients with endometrial cancer (EC).

Methods This multicenter, retrospective study enrolled 892 patients with postoperative pathologically confirmed EC. Preoperative MRI comprised T2-weighted imaging, contrast-enhanced T1-weighted imaging, and apparent diffusion coefficient maps, were analyzed. Regions of interest (ROIs) were manually delineated for 2D and 3D analyses. RFs were extracted using PyRadiomics, and DFs were obtained using pretrained VGG 11, ResNet 101, and DenseNet 121 architectures. Five single-modality models (2D-RF, 3D-RF, VGG11-DF, ResNet101-DF, and DenseNet121-DF) were developed. In addition, the integration of RFs and DFs were explored to construct combined models. Models were trained in a training cohort ($n=378$) and evaluated in both internal ($n=160$) and external ($n=354$) validation cohorts. Model performance was evaluated by the area under the receiver operating characteristic curve (AUC).

Results In the training cohort, the 2D-RF and 3D-RF models showed comparable performance for LVSI prediction (AUC: 0.775 vs. 0.772, $P=0.89$). Among the deep learning models, DenseNet121-DF achieved the highest AUC (0.757), which was significantly higher than ResNet-101-DF (AUC: 0.671; $P=0.01$) and not statistically different from VGG11-DF (AUC: 0.720, $P=0.20$). The optimal combined model, integrating features from 2D-RF and DenseNet121-DF, yielded the highest performance in the training cohort (AUC: 0.796). These findings were confirmed in both the internal and external validation cohorts.

Conclusions A multimodal MRI-based model integrating both RFs and DFs achieved superior performance for noninvasive prediction of LVSI in patients with EC. This approach holds potential to enhance preoperative risk stratification and guide personalized treatment planning.

†Hong-Jian Luo and Jin Cheng contributed equally to this work.

*Correspondence:
Xiao-Li Song
songxiaoli99@126.com

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

Keywords Magnetic resonance imaging, Radiomics, Deep learning, Endometrial cancer, Lymphovascular space invasion

Introduction

Endometrial cancer (EC) is the second most common gynecologic malignancy in China, with an increasing incidence [1]. Contemporary treatment strategies increasingly emphasize personalized care. The 2021 joint guidelines from the European Society of Gynecological Oncology (ESGO), the European Society for Radiotherapy & Oncology (ESTRO), and the European Society of Pathology (ESP) recommended integrating conventional histopathological factors, such as lymphovascular space invasion (LVSI), with molecular classifiers to refine prognostic risk stratification for EC [2]. LVSI represents a critical step in lymphatic dissemination and is an established independent adverse prognosis factor [3, 4]. Accurate determination of LVSI status is crucial for surgical planning, particularly regarding the necessity of lymphadenectomy. The role of lymphadenectomy in the management of EC remains a debated topic. Current guidelines indicate that it can be omitted for patients with low-risk EC without LVSI, thereby reducing complications such as lower limb lymphedema and surgical morbidity [3, 5]. However, definitive assessment of LVSI relies on postoperative histopathology, as intraoperative frozen section analysis is prone to diagnostic errors [6, 7]. Therefore, there is an urgent need for a noninvasive and reliable method to predict LVSI preoperatively in EC.

Magnetic resonance imaging (MRI) plays an important role in the noninvasive management of EC. However, the diagnosis accuracy of conventional MRI for the preoperative assessment of LVSI remains limited. Recently, artificial intelligence has shown significantly potential to advance precision oncology, particularly through the application of radiomics and deep learning [8, 9]. Radiomic, the high-throughput extraction of quantitative features from medical images, has shown promise in correlating MRI-based features with adverse prognostic factors in EC, including deep myometrial invasion (DMI), LVSI, and high histological grade [10, 11]. Nonetheless, the performance of radiomics model is sensitive to methodological choices, including segmentation approach, feature extraction settings, and preprocessing steps [12]. To address these challenges and improve reproducibility, recent studies increasingly recommend that all parameter settings be reported at each step of the radiomics workflow [13–15]. Moreover, conventional handcrafted radiomics features (RFs) which are derived from predefined mathematical formulations, may not fully capture the complex, hierarchical patterns that reflect underlying tumor biology [16].

Deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated a capacity for complex image analysis, often surpassing models based on handcrafted RFs when trained on large scaled, well-annotated datasets [17, 18]. A key advantage of CNNs is their ability to learn hierarchical, high-dimensional representations directly from raw images, thereby reducing reliance on manual feature engineering [19]. Their architectures capture both local and global contextual information, yielding more discriminative and robust features sets [20]. Consequently, advanced CNN architectures such as VGG, ResNet, and DenseNet have achieved promising results across diverse medical image applications [21, 22]. However, their application to preoperative prediction of LVSI in EC remains underexplored. Although prior studies have applied radiomics for this task, many have been constrained by relatively small sample sizes, which may limit the generalizability of their findings [23, 24]. Furthermore, few investigations have explored the potential synergy of integrating handcrafted RFs with deep learning features (DFs). We hypothesize that these feature types are complementary, with RFs adept at characterizing well-defined textural patterns while DFs excel at learning abstract, data-drive representations of tumor heterogeneity. Therefore, this study aimed to develop and validate a multimodal model that integrates both RFs and DFs from multisequence MRI to provide a reliable, noninvasive tool for preoperative prediction of LVSI in patients with EC.

Methods

Patients

This retrospective, multicenter study was conducted in accordance with the Declaration of Helsinki and received approval from the institutional review boards of all participating centers. Informed consent was waived given retrospective nature of design. A total of 1265 patients was initially identified from three institutions between January 2013 and July 2024. Inclusion criteria were: (1) availability of a preoperative MRI examination comprising T2-weighted imaging (T2WI), fat saturation contrast-enhanced T1-weighted imaging (CET1WI), and apparent diffusion coefficient (ADC); and (2) definitive postoperative histopathological confirmation of EC. Exclusion criteria were as follows: (1) receipt of therapy prior to MRI ($n = 97$); (2) history of another tumor ($n = 16$); (3) missing LVSI status ($n = 149$); (4) primary tumor diameters less than 0.5 cm ($n = 42$); and (5) poor image quality ($n = 69$). Finally, 892 patients were enrolled: center I ($n = 538$), center II ($n = 289$), and center III ($n = 65$).

For model development and validation, data from Center I was randomly partitioned into a training cohort (282 LVSI-negative, 96 LVSI-positive) and an internal validation cohort (120 LVSI-negative, 40 LVSI-positive) at a 7:3 ratio. Data from Center II ($n = 289$) and Center III ($n = 65$) were pooled to form an independent external validation cohort ($n = 354$; 269 LVSI-negative, 85 LVSI-positive). Patient enrollment and cohort allocation are summarized in Fig. 1.

Clinicopathological characteristics, including age, histological subtype, DMI, tumor stage, and LVSI status, were extracted from medical records. Histological subtypes were categorized as endometrioid carcinoma and non-endometrioid carcinoma (serous carcinoma, clear cell carcinoma, mixed carcinoma, undifferentiated carcinoma and dedifferentiated carcinoma). Tumors were staged according to the 2009 International Federation of Gynecology and Obstetrics (FIGO) stage system and classified as early stage (I-II) or advanced stage (III-IV).

MRI acquisition and imaging segmentation

All patients underwent preoperative pelvic contrast-enhanced MRI. The standardized imaging protocols included axial T2WI, diffusion weighted imaging (DWI), and fat saturation CET1WI acquired in the delayed phase following intravenous administration of a gadolinium-based contrast agent (gadopentetate dimeglumine and gadodiamide). Apparent diffusion coefficient (ADC) maps were automatically generated from the DWI data using a monoexponential model. Detailed information on MRI scanners and acquisition parameters is provided in Supplementary Material S1.

Tumors regions of interests (ROIs) were manually delineated on T2WI, CET1WI, and DWI using ITK-SNAP software (version 3.8.0; https://www.itksnap.org) by a radiologist with 8 years of experience in pelvic imaging. For each patient, both three-dimensional (3D) ROIs covering the entire tumor volume, and two-dimensional (2D) ROIs drawn on the slice with the largest tumor area were generated (Fig. 2). Interobserver variability was assessed in a randomly selected subset of 100 patients, which underwent independent segmentation by a second radiologist with 5 years of experience. Both readers were blinded to histopathology results.

Imaging preprocessing and radiomics feature extraction

Image preprocessing and radiomics features (RFs) extraction were performed using PyRadiomics package (version 3.1.0) in Python (version 3.9) [25] in accordance with Image Biomarker Standardization Initiative (IBSI) guidelines [15]. Prior to feature extraction, the following preprocessing steps were applied: (1) gray-level normalization using z-score transformation; (2) voxel resampling and rescaling using B-spline interpolation (1 mm × 1 mm × 1 mm for T2WI and CET1WI, 1.5 mm × 1.5 mm × 1.5 mm for ADC maps); and (3) gray-level discretization with a fixed bin width of 5. Preprocessing parameters were chosen based on preliminary experiments

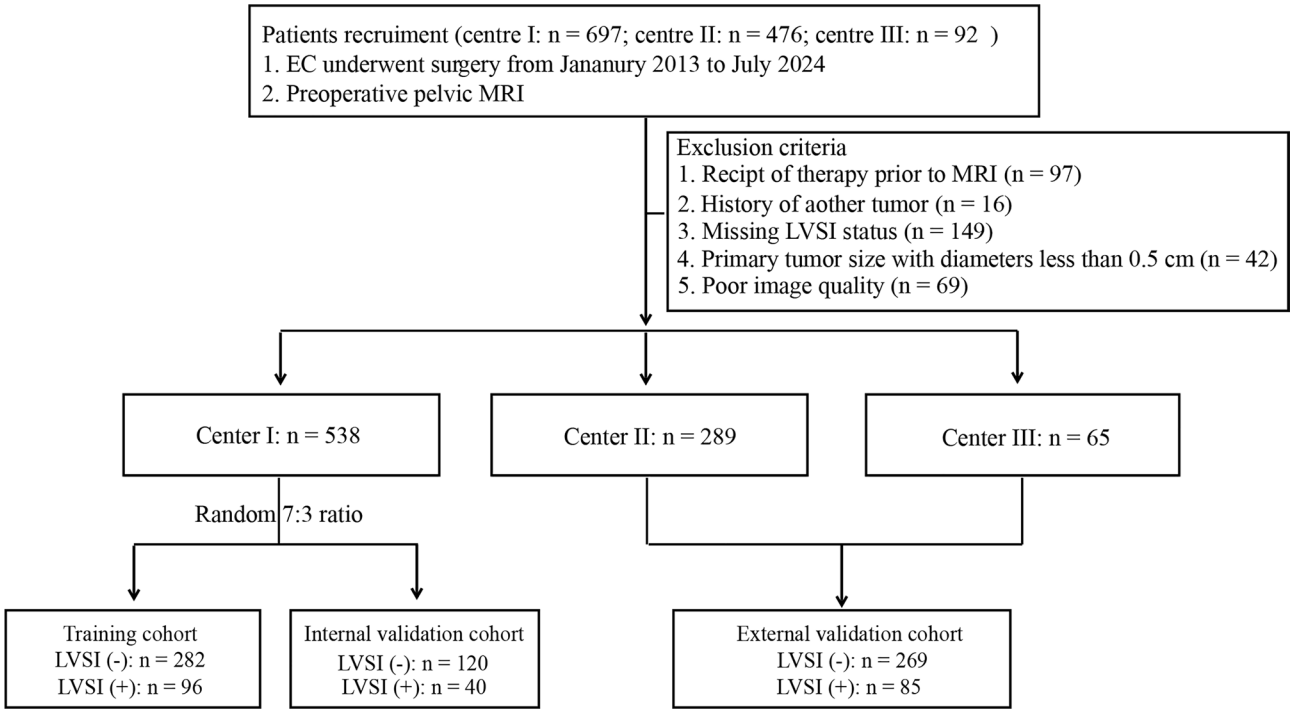

Fig. 1 Patient selection process and study design. LVSI: lymphovascular space invasion

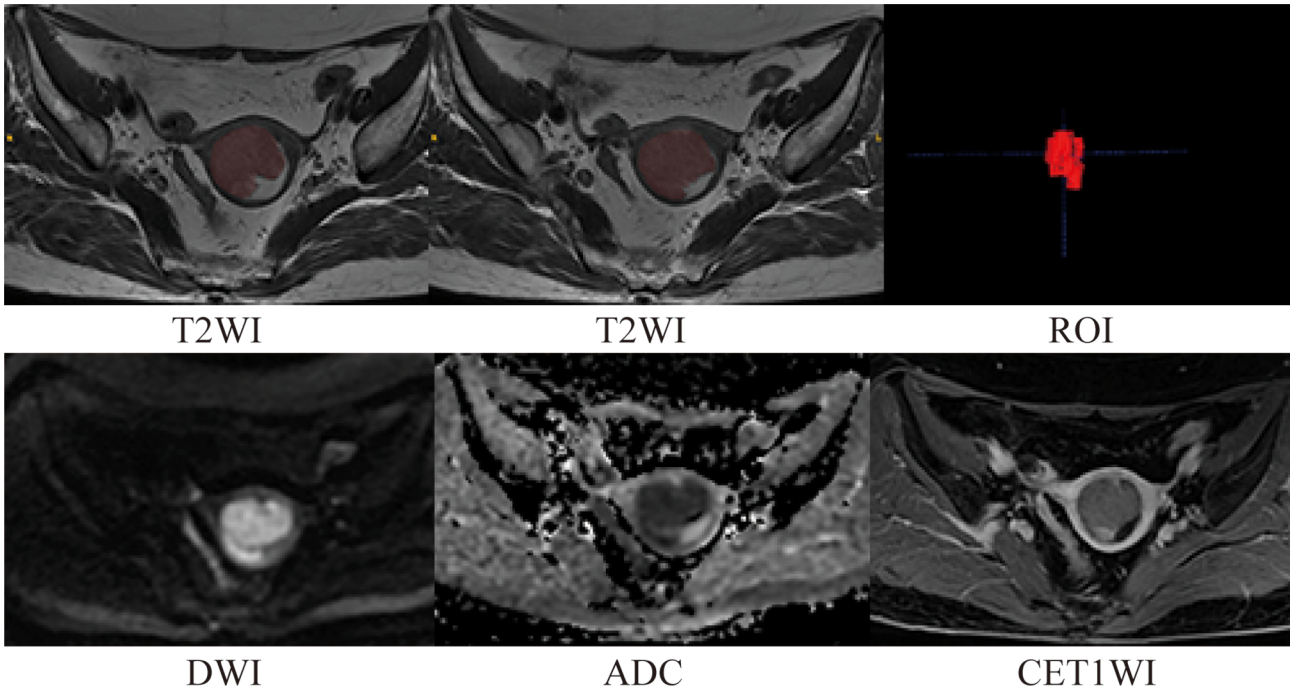

Fig. 2 Manual segmentation of the lesion in a patient with endometrial cancer. The top row illustrates the tumor boundary is delineated on an axial T2-weighted image (T2WI) and the corresponding 3D region of interest (ROI). The bottom row displays the corresponding images from diffusion-weighted image (DWI, $b = 1000$ s/mm^2), apparent diffusion coefficient (ADC) map, and contrast-enhanced T1-weighted image (CE-T1WI)

and established literature recommendations to optimize feature stability and discriminative power. Following preprocessing, seven classes of RFs were extracted: first-order features, shape features, gray-level cooccurrence matrix (GLCM) features, gray-level run-length matrix (GLRLM) features, gray-level size zone matrix (GLSZM) features, gray-level dependence matrix (GLDM) features, and neighboring gray-tone difference matrix (NGTDM) features. This process yielded 3933 2D RFs and 5064 3D RFs for each patient. Further details of RFs extraction are provided in Supplementary Material S2.

Deep learning feature extraction

DFs were extracted using a transfer learning approach with three pretrained CNN architectures: VGG11, ResNet101, and DenseNet121. To balance feature representativeness with computational efficiency, 2D ROIs were used as the inputs to the model. Images were preprocessing by resizing them to 224×224 pixels and normalizing pixel intensities values (zero mean, unit variance) to meet the input requirements of the networks. For each CNN, the final fully connected layers were removed, and the remaining convolutional backbone was used as a fixed feature extractor. Features were obtained from the output of the last convolutional block in each architecture and converted into a one-dimensional feature via global average pooling. All implementations were performed in PyTorch (version 1.9.0). All DFs were z-score standardized before model development. Further details on the network configurations and implementation are provided in Supplementary Material S3. This multi-architecture strategy was intended to capture complementary abstractions across diverse architectures.

Feature selection and model building

A four-step features selection approach was performed in the training cohort to identify the optimal predictive features and prevent model overfiting. First, each feature was assessed using univariable logistic regression, and only those with a statistically significant association with LVSI ($P < 0.05$) were retained. Second, to mitigate multicollinearity, Pearson correlation analysis was performed, and highly correlated features ($r > 0.9$) were removed. Third, the remaining features were entered into a least absolute shrinkage and selection operator (LASSO) method. The optimal regularization parameter (λ) was determined by 10-fold cross-validation by using minimized mean cross-validation error. Finally, backward stepwise elimination guided by the Akaike's information criterion (AIC) was applied. This iterative process was terminated when the removal of any additional feature failed to further reduce the AIC, yielding a parsimonious feature set for model development.

Using the selected features, six multivariable logistic regression models were constructed. The first two models were based on RFs: a 2D-RF model (Model 1) and a

3D-RF model (Model 2), derived from 2D and 3D ROIs, respectively. We then developed three separate models using deep features (DFs) extracted from pre-trained networks: a VGG11-DF model (Model 3), a ResNet101-DF model (Model 4), and a DenseNet121-DF model (Model 5). Finally, a combined model (Model 6) was constructed by integration of features from the RF models and the DF models.

Statistical analysis

All statistical analyses were performed using the R software (version 4.4.0, https://www.rproject.org). Baseline characteristics were compared using the Kruskal-Wallis H test for continuous variables and the chi-square test for categorical variables. Interobserver agreement for RFs was assessed using the intraclass correlation coefficients (ICC). Model discrimination was quantified by the area under the receiver operating characteristic curve (AUC), with AUCs compared using DeLong's test. To address class imbalance, the area under the precision-recall curve was also calculated. Accuracy, sensitivity, specificity, and F1 score were reported at an optimal probability threshold determined by Youden's index in the training cohort and then applied to the validation cohort. Model calibration was evaluated with the Hosmer-Lemeshow test. Potential overfitting was examined via bootstrap optimism correction with 1,000 resamples. Multicollinearity was assesses using the variance inflation factor (VIF), with values < 5 considered acceptable. Clinical utility was evaluated by decision curve analysis (DCA), which estimates net benefit across a range of threshold probabilities. Two-sided $P < 0.05$ was considered statistically significance.

Results

Clinicopathological characteristics of EC patients

A total of 892 patients diagnosed with EC were enrolled in this retrospective multicenter study (538 from Center I, 289 from Center II, and 65 from Center III). Patients from Center I were randomly assigned to training ($n = 378$) and internal validation ($n = 160$) cohorts. Patients from Center II and Center III formed an independent external validation cohort ($n = 354$). The prevalence of LVSI was similar across the training cohort (25.4%), the internal validation cohort (25.0%), external validation cohort (24.0%). Endometrioid carcinoma was the predominant histological subtype across all cohorts, ranging from 87.3% to 90.6%. No significant differences in baseline characteristics were observed among the cohorts (all $P > 0.05$), as detailed in Table 1.

Feature selection

A multistep feature selection process was applied to identify the optimal features for predicting LVSI status in EC. An initial univariate screen ($P < 0.05$) identified a large pool of candidate features, ranging from 929 to 1,720 across the different input sets. Highly correlated features (Pearson's $r > 0.9$) were then removed, followed by the application of LASSO regression. Backward stepwise logistic regression guided by AIC produced the final feature sets for the five model: 3 features in model 1(2D-RF), 4 features in Model 2 (3D-RF), 5 features in Model 3 (VGG11-DF), 6 features in Model 4 (ResNet101-DF), and 9 features in Model 5 (DenseNet121-DF). All retained features demonstrated good to excellent interobserver agreement via ICC analysis, indicating robustness to segmentation variability. Detailed features for each model are provided in Table 2, with corresponding ICC values in Supplementary Material S4.

Table 1 Baseline clinicopathological characteristics of the patients with endometrial cancer in the training and validation cohorts

	Training cohort ($n = 378$)	Internal validation cohort ($n = 160$)	External validation cohorts ($n = 354$)	*P*-value
Age [median (IQR)]	57.0 (52.0, 63.0)	58.0 (52.0, 63.0)	57.0 (52.0, 63.0)	0.544
Pathological type [n (%)]				0.474
Endometrioid carcinoma	338 (89.4%)	145 (90.6%)	309 (87.3%)	
Non-endometrioid carcinoma	40 (10.6%)	15 (9.4%)	45 (12.7%)	
FIGO stage [n (%)]				0.072
Early-stage	338 (89.4%)	143 (89.4%)	298 (84.2%)	
Late-stage	40 (10.6%)	17 (10.6%)	56 (15.8%)	
DMI [n (%)]				0.225
-	278 (73.5%)	110 (68.8%)	269 (76.0%)	
+	100 (26.5%)	50 (31.2%)	85 (24.0%)	
LVSI status [n (%)]				0.908
-	282 (74.6%)	120 (75.0%)	269 (76.0%)	
+	96 (25.4%)	40 (25.0%)	85 (24.0%)	

IQR: interquartile range; FIGO: International Federation of Gynecology and Obstetrics; DMI: deep myometrial invasion; LVSI: lymphovascular space invasion

Table 2 Details of retained features in each model

Model	Feature name	Coefficient	*P*	Odd Ratio
2D-RF	T2W_log.sigma.3.0.mm.3D_firstorder_10Percentile	1.148	< 0.001	3.150 (2.212–4.657)
	T2W_wavelet.LH_ngtdm_Contrast	0.335	0.024	1.398 (1.038–1.903)
	T2W_exponential_glrlm_RunEntropy	0.888	< 0.001	2.425 (1.804–3.328)
3D-RF	ADC_original_firstorder_Kurtosis	0.370	0.007	1.448 (1.126–1.942)
	CET1W_wavelet.LHL_glcm_Idn	0.289	0.124	1.335 (0.936–1.954)
	T2W_log.sigma.3.0.mm.3D_firstorder_10Percentile	0.816	< 0.001	2.261 (1.636–3.223)
	T2W_exponential_glrlm_RunEntropy	0.357	0.034	1.430 (1.033–2.004)
VGG11-DF	CET1W_F1813	0.241	0.054	1.272 (0.995–1.628)
	CET1W_F2153	0.407	0.001	1.502 (1.195–1.898)
	T2W_F175	-0.327	0.044	0.721 (0.516–0.978)
	T2W_F1618	0.206	0.076	1.229 (0.978–1.549)
	T2W_F2287	-0.284	0.132	0.753 (0.503–1.058)
ResNet101-DF	ADC_F365	-0.255	0.068	0.775 (0.584–1.011)
	CET1W_F862	0.337	0.028	1.400 (1.069–1.930)
	CET1W_F1075	-0.338	0.038	0.713 (0.507–0.965)
	T2W_F69	0.238	0.046	1.269 (1.010–1.626)
	T2W_F273	-0.321	0.069	0.725 (0.499–1.001)
	T2W_F1320	0.218	0.061	1.244 (0.985–1.564)
DenseNet121-DF	ADC_F756	0.260	0.048	1.297 (1.003–1.681)
	ADC_F1304	0.213	0.073	1.237 (0.984–1.577)
	ADC_F1628	0.267	0.030	1.305 (1.034–1.680)
	CET1W_F74	0.231	0.085	1.260 (0.968–1.642)
	T2W_F59	0.239	0.077	1.270 (0.975–1.659)
	T2W_F824	0.314	0.022	1.369 (1.048–1.798)
	T2W_F980	0.231	0.082	1.260 (0.972–1.639)
	T2W_F1124	-0.270	0.059	0.763 (0.574–1.006)
	T2W_F1584	-0.226	0.112	0.798 (0.600–1.050)

T2WI: T2-weighted imaging; ADC: apparent diffusion coefficient; CET1WI: contrast-enhanced T1-weighted imaging

Predictive performance of the models

In the training cohort, the 2D-RF model achieved an AUC of 0.775, comparable to the 3D-RF model (AUC = 0.772, $P = 0.89$). Among the deep learning models, the DenseNet121-DF model (AUC = 0.757) outperformed the VGG11-DF and ResNet101-DF models. These trends were confirmed in the internal and external validation cohorts.

To further enhance discrimination, combined models were constructed by integrating selected features from the RF and DF models. The optimal combined model (Model 6), which integrating features from the 2D-RF and DenseNet121-DF models, achieved the highest overall performance in the training cohort (AUC: 0.796). The Model 6 outperformed the Model 5 (AUC: 0.796 vs. 0.757, $P = 0.01$) and Model 1 (AUC: 0.796 vs. 0.775, $P = 0.16$). Model 6 maintained robust discrimination in the internal and external validation cohorts (AUC: 0.751 and 0.765, respectively). Detailed metrics (accuracy, sensitivity, specificity, and AUC) for all models across cohorts are summarized in Table 3; Fig. 3, with alternative fusion configurations is summarized in Supplementary Material S5.

Model validation, calibration, and clinical utility

The stability of the final combined model was assessed. Bootstrap resampling with 1,000 iterations demonstrated minimal optimism in performance estimates, indicating a low risk of overfitting (Supplementary Material S6). The precision-recall curve demonstrated consistently high precision across a broad range of recall values, supporting the effectiveness of the model under class imbalance (Supplementary Material S7). To evaluate clinical applicability, we performed a DCA. The results showed a greater net clinical benefit for LVSI status prediction, particularly for threshold probabilities of 0.20–0.50, a clinically relevant range for decision-making (Fig. 4).

The model also exhibited excellent calibration across all cohorts, with calibration curves showing close agreement between predicted probabilities and observed LVSI frequencies (Fig. 4). The Hosmer-Lemeshow test indicated no significant lack of fit ($P > 0.05$ for all cohorts). Detailed calibration metrics, including calibration intercept (α), calibration slope (β), and the Brier score, are summarized in Supplementary Material S8.

Table 3 Performance of models for the prediction of lymphovascular space invasion in endometrial cancer in the training and validation cohorts

	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPV (95% CI)	NPV (95% CI)	F1
Training cohort							
2D-RF model	0.775(0.722–0.827)	0.659(0.609–0.706)	0.823(0.698–0.907)	0.603(0.443–0.681)	0.414(0.374–0.437)	0.909(0.880–0.919)	0.551
3D-RF model	0.772(0.718–0.825)	0.741(0.693–0.784)	0.656(0.479–0.729)	0.770(0.603–0.823)	0.492(0.414–0.519)	0.868(0.837–0.875)	0.565
VGG11-DF model	0.720(0.657–0.782)	0.709(0.660–0.754)	0.635(0.500–0.740)	0.734(0.542–0.809)	0.449(0.390–0.486)	0.855(0.814–0.867)	0.526
ResNet101-DF model	0.671(0.605–0.736)	0.743(0.696–0.787)	0.448(0.323–0.521)	0.844(0.713–0.904)	0.494(0.413–0.532)	0.818(0.791–0.828)	0.470
DenseNet121-DF model	0.757(0.699–0.815)	0.751(0.705–0.794)	0.604(0.489–0.688)	0.801(0.652–0.890)	0.509(0.456–0.541)	0.856(0.829–0.869)	0.552
Combined model	0.796(0.745–0.846)	0.746(0.699–0.789)	0.708(0.583–0.781)	0.759(0.610–0.830)	0.500(0.452–0.524)	0.884(0.860–0.893)	0.586
Internal validation cohort							
2D-RF model	0.738(0.643–0.833)	0.619(0.539–0.694)	0.750(0.624-0.900)	0.575(0.250–0.825)	0.370(0.329–0.414)	0.873(0.750–0.908)	0.500
3D-RF model	0.724(0.635–0.813)	0.706(0.629–0.776)	0.625(0.474-0.800)	0.733(0.508–0.867)	0.439(0.372-0.500)	0.854(0.803–0.874)	0.515
VGG11-DF model	0.712(0.625–0.799)	0.688(0.610–0.758)	0.575(0.349–0.775)	0.725(0.558–0.833)	0.411(0.297–0.484)	0.837(0.797–0.855)	0.479
ResNet101-DF model	0.723(0.638–0.807)	0.738(0.662–0.804)	0.475(0.225–0.675)	0.825(0.675–0.908)	0.475(0.300-0.562)	0.825(0.794–0.838)	0.475
DenseNet121-DF model	0.743(0.656–0.830)	0.681(0.603–0.753)	0.600(0.450–0.800)	0.708(0.592–0.917)	0.407(0.340–0.478)	0.842(0.816–0.873)	0.485
Combined model	0.751(0.659–0.843)	0.700(0.623–0.770)	0.675(0.525–0.826)	0.708(0.533–0.917)	0.435(0.375–0.485)	0.867(0.831–0.894)	0.529
External validation cohort							
2D-RF model	0.725(0.664–0.785)	0.565(0.512–0.617)	0.847(0.765–0.906)	0.476(0.342–0.606)	0.338(0.316–0.353)	0.908(0.876–0.926)	0.483
3D-RF model	0.730(0.670–0.791)	0.698(0.647–0.745)	0.659(0.506–0.776)	0.710(0.587–0.807)	0.418(0.355–0.458)	0.868(0.845–0.882)	0.511
VGG11-DF model	0.720(0.659–0.782)	0.647(0.595–0.697)	0.694(0.600-0.812)	0.632(0.520–0.732)	0.373(0.340–0.411)	0.867(0.843–0.883)	0.486
ResNet101-DF model	0.698(0.636–0.761)	0.726(0.676–0.772)	0.435(0.294–0.565)	0.818(0.699–0.877)	0.430(0.338–0.495)	0.821(0.797–0.831)	0.433
DenseNet121-DF model	0.757(0.699–0.815)	0.706(0.656–0.753)	0.671(0.576–0.753)	0.717(0.602–0.840)	0.429(0.392–0.457)	0.873(0.853–0.890)	0.523
Combined model	0.765(0.707–0.823)	0.647(0.595–0.697)	0.788(0.706–0.882)	0.602(0.433–0.751)	0.385(0.359–0.412)	0.900(0.866–0.918)	0.517

AUC: area under the curve; ACC: accuracy; SEN: sensitivity; SPE: specificity; PPV: positive prediction rate; NPV: negative prediction rate; 95% CI: 95% confidence interval

Discussion

In this multicenter study, a combined model that integrating RFs and DFs derived from preoperative MRI was developed and validated for the noninvasive prediction of LVSI in EC. Our results demonstrated that the 2D-RF model performed comparably to the 3D-RF model, suggesting that the less labor-intensive 2D segmentation approach is sufficient for this task. Among the evaluated CNN architectures, DenseNet-121 yield the most informative DFs. Critically, the combined RF-DF model achieved superior predictive performance, supporting the hypothesis that handcrafted and deep-learned features provided complementarity information for assessing LVSI status in EC.

Accurate preoperative identification of LVSI is crucial for guiding surgical planning and adjuvant therapy in EC. In our cohort, the 2D-RF model achieved an AUC of 0.775, which is comparable to the performance reported in previous MRI-based radiomic studies for LVSI prediction. Ueno et al. [26] developed a random forest classifier that achieved an AUC of 0.80, whereas Luo et al. [10] reported AUCs of 0.820 and 0.807 in its training and test cohorts, respectively. Minor performance difference across studies may be attributed to variations in patient populations, imaging protocols, and cohort sizes. In addition, our study design addressed the trade-off between comprehensive tumor characterization and clinical pragmatism in selecting between 2D and 3D analysis. Consistent with prior studies in gastric cancer [27], lung adenocarcinoma [28], and EC [29], which also found limited benefit from 3D analysis. While 3D segmentation theoretically offers a more comprehensive assessment of intratumoral heterogeneity, its time-intensive nature hampered its clinical application. More importantly, our multicenter dataset comprised anisotropic 2D slices with interslice gaps rather than contiguous, isotropic 3D

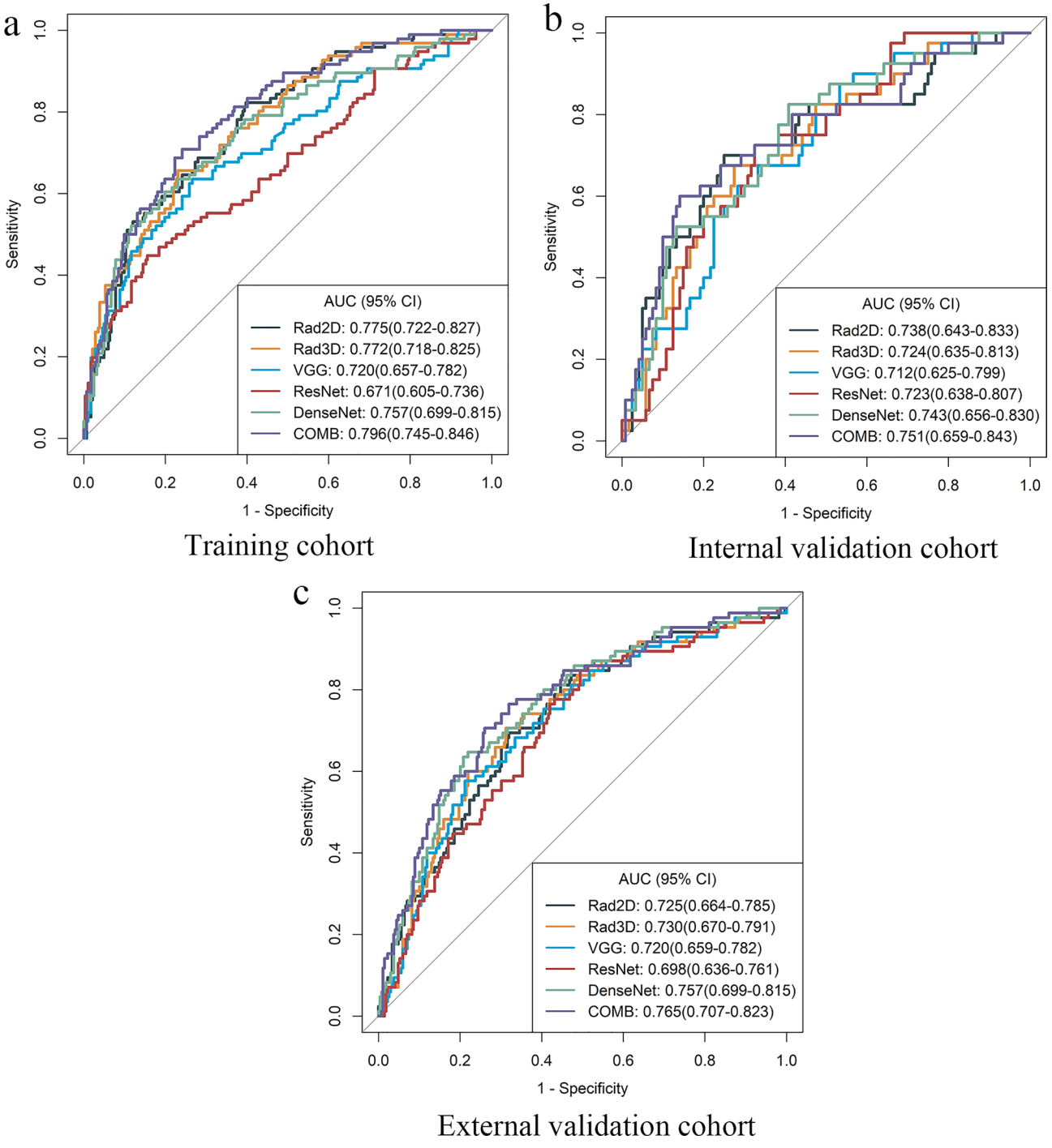

Fig. 3 Receiver operating characteristic curves of the predictive models for lymphovascular space invasion status in patients with endometrial cancer, shown for (**a**) training cohort, (**b**) internal validation cohort, (**c**) external validation cohort

volumes. A 3D approach would therefore have required through-plane interpolation to synthesize missing data, introducing the risk that the models learn algorithmic artifacts rather than true tumor biology [30]. Therefore, our adoption of a 2D approach was not only clinically pragmatic but also methodologically robust, as it relied exclusively on features from directly acquired data and avoided potential interpolation-related confounders.

Unlike handcrafted radiomics, deep learning models can automatically learn hierarchical features from raw image, thereby reducing reliance on manual feature engineering. We extended our 2D framework to the deep learning approach, a decision grounded in the

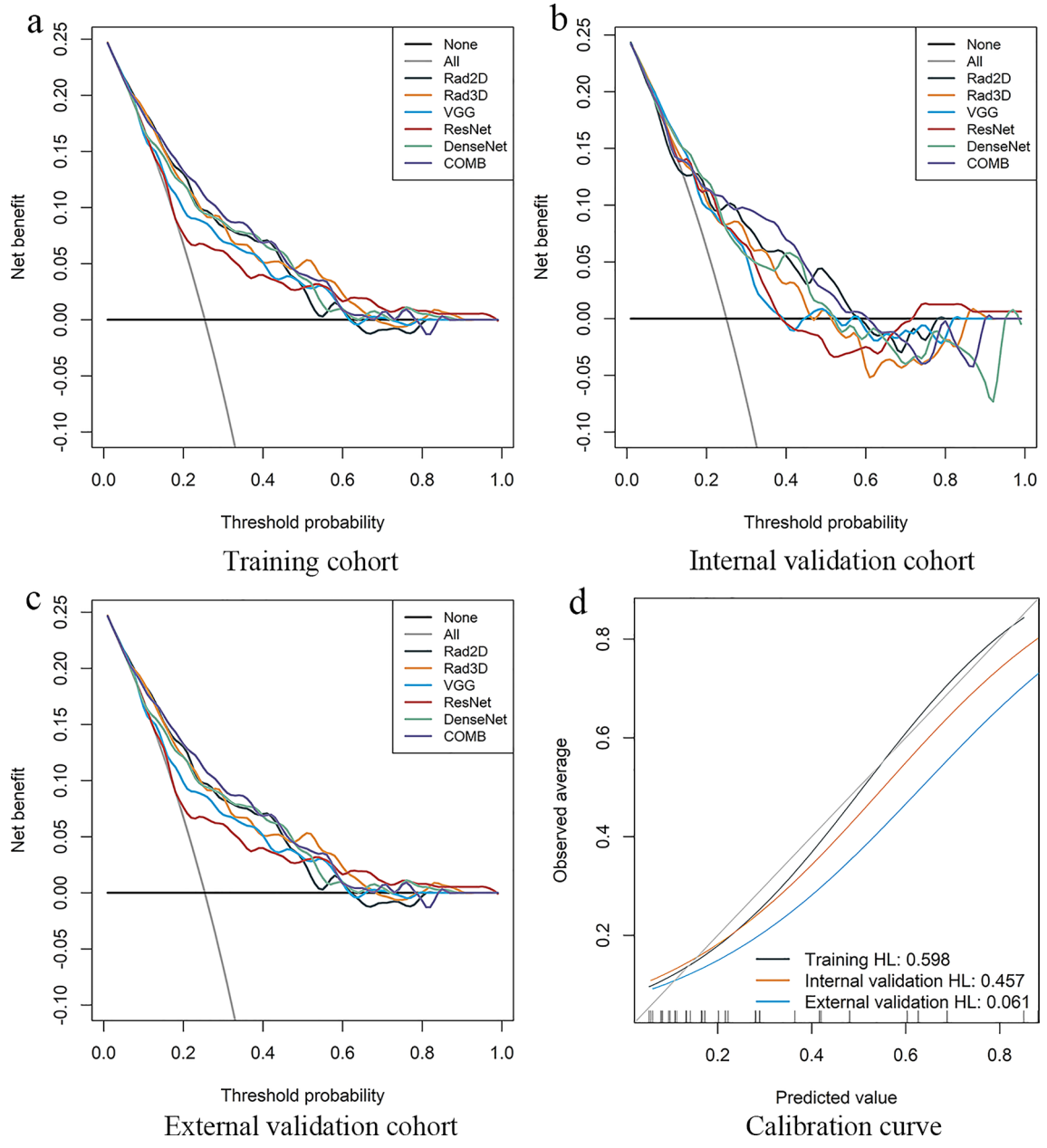

Fig. 4 Decision curve analysis of the models for predicting lymphovascular space invasion status in the (**a**) training cohort, (**b**) internal validation cohort, (**c**) external validation cohorts. The y-axis represents the net benefit, while the x-axis represents the threshold probability. Calibration curves (**d**) for the combined model across all three cohorts, plotting the predicted probability against the actual observed frequency of LVSI

critical principle of avoiding interpolation artifacts from anisotropic source data. Beyond this primary rationale, a 2D CNN framework offers distinct practical advantages, including reduced computational burden and a lower requirement for annotated samples compared to 3D CNNs, which enhances the feasibility of large-scale, multicenter research [31, 32]. To identify the optimal architecture, we evaluated three CNN architectures of VGG11, ResNet101, and DenseNet121. VGG11 served as a shallow baseline, ResNet101 assessed the utility of a very deep network with residual connections, and DenseNet121 was selected for its characteristic dense connectivity. DenseNet121 demonstrated superior predictive performance, an outcome attributable to its unique architecture, which promotes extensive feature reuse and efficient gradient flow, thereby mitigating the vanishing gradient during training [33]. This finding aligns with the documented successes of DenseNet121 in other complex medical imaging tasks, such as staging peritoneal metastasis [34] and classifying breast tumors [35].

Our findings support the value of a multimodal strategy, a well-established principle in personalized medicine [36, 37]. The combined RF-DF model outperformed models using either feature type alone, indicating that RFs and DFs capture synergistic and complementary biological information. This enhanced discrimination was accompanied by excellent calibration across all cohorts (Hosmer-Lemeshow test, all P >0.05), supporting the model's reliability for clinical application. Furthermore, DCA confirmed that the combined model provides a consistent net clinical benefit over treat-all or treat-none strategies. This benefit was most pronounced across a clinically relevant threshold probability range of 0.20-0.50, with the greatest separation around 0.25-0.45. This threshold range is highly pertinent to preoperative EC management, where LVSI risk assessment directly informs decision regarding the extent of lymphadenectomy and the need for adjuvant therapy. Notably, the optimal operating point derived from the training cohort via Youden's index falls within this window, reinforcing the model's practical applicability for preoperative risk stratification.

This study has several limitations. First, the retrospective design necessitated complete-case analysis, which may have introduced selection bias. Prospective studies with prespecified data collection and robust handling missing data, such as multiple imputation, are warranted to validate and generalize our findings. Second, although the retained features showed high inter-rater reliability, reliance on manual tumor segmentation is labor-intensive and remains a barrier to routine clinical adoption, highlighting the need for robust, automated segmentation algorithms. Third, the cohort was composed of patients with primary EC. The generalizability to recurrent or metastatic disease is uncertain and merits dedicated evaluation. Finally, the deep-learning components were based on CNNs pretrained on ImageNet. While this transfer learning approach is effective, the domain gap between natural and medical images may limit features relevance. Pretrained on large, domain-specific medical datasets could further improve performance.

Conclusions

In this multicenter study, we developed and validated an integrated model that combines DFs and RFs derived from preoperative MRI for noninvasive prediction of LVSI in EC. This multimodal approach improved performance over either feature type alone and has the potential to enhance preoperative risk stratification and guide individualized surgical planning. To facilitate clinical translation, future work should prioritize prospective multicenter validation, integrate automated segmentation, and rigorously evaluate performance in larger, more diverse patient populations, including those with recurrent or metastatic disease, to establish clinical utility.

Abbreviations

EC	Endometrial cancer
LVSI	Lymphovascular space invasion
RF	Radiomics features
DF	Deep learning features
ADC	Apparent diffusion coefficient
LASSO	Least absolute shrinkage and selection operator
AIC	Akaike's information criterion
DCA	Decision curve analysis
GLCM	Gray-level cooccurrence matrix
GLRLM	Gray-level run length matrix
GLDM	Gray-level dependence matrix
GLSZM	Gray-level size zone matrix
NGTDM	Neighborhood gray-tone difference matrix

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-02091-4.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

X.L.S., H.J.L., J.C., and K.W. conceived and designed research; K.W., and J.L.R. performed experiments; J.L.R., L.M.G., and J.L.N. interpreted results of experiments. X.L.S., H.J.L., and J.C. drafted manuscript; X.L.S., H.J.L., J.C., and J.L.R. edited and revised manuscript; All authors approved final version of manuscript.

Funding

Not applicable.

Data availability

The analysis pipeline was implemented using open-source tools. Radiomic features were extracted with PyRadiomics (https://github.com/AIM-Harvard/pyradiomics). Deep learning models were initialized with weights from the pretrained-models.pytorch repository (https://github.com/Cadene/pretrained-models.pytorch). The datasets generated and/or analyzed in this study, including the finalized feature sets and analysis code, are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the institutional review boards of the Third Affiliated Hospital of Zunyi Medical University (No.2024-1-491), Peking University People's Hospital (No.2025PHB057-001), and Second Hospital of Shanxi Medical University (No.2024YX115) in accordance with the Declaration of Helsinki. The requirement for written informed consent was waived due to the retrospective nature of the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Radiology, The Third Affiliated Hospital of Zunyi Medical University (The First People's Hospital of Zunyi), Zunyi, Guizhou 563000, China

[2]Department of Radiology, Peking University People's Hospital, 11 Xizhimen Nandajie, Xicheng District, Beijing 100044, China
[3]GE Healthcare, Beijing 100176, China
[4]Department of Radiology, Second Hospital of Shanxi Medical University, Taiyuan, Shanxi 030001, China

Received: 15 May 2025 / Accepted: 25 November 2025
Published online: 08 December 2025

References

1. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent. 2024;4:47–53.
2. Concin N, Matias-Guiu X, Vergote I, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. Int J Gynecol Cancer. 2021;31:12–39.
3. Harris KL, Maurer KA, Jarboe E, et al. LVSI positive and NX in early endometrial cancer: surgical restaging (and no further treatment if N0), or adjuvant ERT? Gynecol Oncol. 2020;156:243–50.
4. Raffone A, Travaglino A, Raimondo D, et al. Lymphovascular space invasion in endometrial carcinoma: A prognostic factor independent from molecular signature. Gynecol Oncol. 2022;165:192–7.
5. Colombo N, Creutzberg C, Amant F, et al. ESMO-ESGO-ESTRO consensus conference on endometrial cancer: diagnosis, treatment and follow-up. Ann Oncol. 2016;27:16–41.
6. Berek JS, Matias-Guiu X, Creutzberg C, et al. FIGO staging of endometrial cancer: 2023. J Gynecol Oncol. 2023;34:e85.
7. Oliver-Perez MR, Magriña J, Villalain-Gonzalez C, et al. Lymphovascular space invasion in endometrial carcinoma: tumor size and location matter. Surg Oncol. 2021;37:101541.
8. Huisman M, Akinci D'Antonoli T. What a radiologist needs to know about Radiomics, Standardization, and reproducibility. Radiology. 2024;310:e232459.
9. Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022;79:102444.
10. Luo Y, Mei D, Gong J, et al. Multiparametric MRI-Based radiomics nomogram for predicting lymphovascular space invasion in endometrial carcinoma. J Magn Reson Imaging. 2020;52:1257–62.
11. Fasmer KE, Hodneland E, Dybvik JA, et al. Whole-Volume tumor MRI radiomics for prognostic modeling in endometrial cancer. J Magn Reson Imaging. 2021;53:928–37.
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than Pictures, they are data. Radiology. 2016;278:563–77.
13. Stefano A. Challenges and limitations in applying radiomics to PET imaging: possible opportunities and avenues for research. Comput Biol Med. 2024;179:108827.
14. Kocak B, Baessler B, Bakas S, et al. CheckList for evaluation of radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII. Insights Imaging. 2023;14:75.
15. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for High-Throughput image-based phenotyping. Radiology. 2020;295:328–38.
16. Yang X, Wu L, Zhao K, et al. Evaluation of human epidermal growth factor receptor 2 status of breast cancer using preoperative multidetector computed tomography with deep learning and handcrafted radiomics features. Chin J Cancer Res. 2020;32:175–85.
17. Ng HW, Guan C. Deep unsupervised representation learning for Feature-Informed EEG domain extraction. IEEE Trans Neural Syst Rehabil Eng. 2023;31:4882–94.
18. Sari CT, Gunduz-Demir C. Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images. IEEE Trans Med Imaging. 2019;38:1139–49.
19. Derry A, Krzywinski M, Altman N. Convolutional neural networks. Nat Methods. 2023;20:1269–70.
20. Soffer S, Ben-Cohen A, Shimon O, et al. Convolutional neural networks for radiologic images: A radiologist's guide. Radiology. 2019;290:590–606.
21. Benyahia S, Meftah B, Lézoray O. Multi-features extraction based on deep learning for skin lesion classification. Tissue Cell. 2022;74:101701.
22. Li Z, Wu F, Hong F, et al. Computer-Aided diagnosis of spinal tuberculosis from CT images based on deep learning with multimodal feature fusion. Front Microbiol. 2022;13:823324.
23. Ma C, Zhao Y, Song Q, et al. Multi-parametric MRI-based radiomics for preoperative prediction of multiple biological characteristics in endometrial cancer. Front Oncol. 2023;13:1280022.
24. Chen J, Wang X, Lv H, et al. Development and external validation of a clinical-radiomics nomogram for preoperative prediction of LVSI status in patients with endometrial carcinoma. J Cancer Res Clin Oncol. 2023;149:13943–53.
25. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to Decode the radiographic phenotype. Cancer Res. 2017;77:e104–7.
26. Ueno Y, Forghani B, Forghani R, et al. Endometrial carcinoma: MR Imaging-based texture model for preoperative risk Stratification-A preliminary analysis. Radiology. 2017;284:748–57.
27. Meng L, Dong D, Chen X, et al. 2D and 3D CT radiomic features performance comparison in characterization of gastric cancer: A Multi-Center study. IEEE J Biomed Health Inf. 2021;25:755–63.
28. Peng Z, Lin Z, He A, et al. Development and validation of a comprehensive model for predicting distant metastasis of solid lung adenocarcinoma: 3D radiomics, 2D radiomics and clinical features. Cancer Manag Res. 2022;14:3437–48.
29. Yan R, Qin S, Xu J, et al. A comparison of 2D and 3D magnetic resonance imaging-based intratumoral and peritumoral radiomics models for the prognostic prediction of endometrial cancer: a pilot study. Cancer Imaging. 2024;24:100.
30. Dong Z, He Y, Qi X, et al. MNet: rethinking 2D/3D networks for anisotropic medical image segmentation. arXiv preprint arXiv:220504846. 2022.
31. Yan M, Zhang X, Zhang B, et al. Deep learning nomogram based on Gd-EOB-DTPA MRI for predicting early recurrence in hepatocellular carcinoma after hepatectomy. Eur Radiol. 2023;33:4949–61.
32. Singh SP, Wang L, Gupta S, et al. 3D deep learning on medical images: A review. Sens (Basel). 2020;20:5097.
33. Peng Y, Zhu W, Chen Z, et al. Automatic staging for retinopathy of prematurity with deep feature fusion and ordinal classification strategy. IEEE Trans Med Imaging. 2021;40:1750–62.
34. Wang J, Hu Y, Xiong H, et al. CT-based deep learning model: a novel approach to the preoperative staging in patients with peritoneal metastasis. Clin Exp Metastasis. 2023;40:493–504.
35. Xu P, Zhao J, Wan M, et al. Classification of multi-feature fusion ultrasound images of breast tumor within category 4 using convolutional neural networks. Med Phys. 2024;51:4243–57.
36. Li M, Fan Y, You H, et al. Dual-Energy CT deep learning radiomics to predict Macrotrabecular-Massive hepatocellular carcinoma. Radiology. 2023;308:e230255.
37. Beuque MPL, Lobbes MBI, van Wijk Y, et al. Combining deep learning and handcrafted radiomics for classification of suspicious lesions on Contrast-enhanced mammograms. Radiology. 2023;307:e221843.

Publisher's note