BMC Cancer

RESEARCH Open Access

Estimation of TP53 mutations for endometrial cancer based on diffusion-weighted imaging deep learning and radiomics features

Lei Shen[1†], Bo Dai[1†], Shewei Dou[1], Fengshan Yan[1], Tianyun Yang[1] and Yaping Wu[1*]

Abstract

Objectives To construct a prediction model based on deep learning (DL) and radiomics features of diffusion weighted imaging (DWI), and clinical variables for evaluating TP53 mutations in endometrial cancer (EC).

Methods DWI and clinical data from 155 EC patients were included in this study, consisting of 80 in the training set, 35 in the test set, and 40 in the external validation set. Radiomics features, convolutional neural network-based DL features, and clinical variables were analyzed. Feature selection was performed using Mann-Whitney U test, LASSO regression, and SelectKBest. Prediction models were established by gaussian process (GP) and decision tree (DT) algorithms and evaluated by the area under the receiver operating characteristic curve (AUC), net reclassification index (NRI), calibration curves, and decision curve analysis (DCA).

Results Compared to the DL ($AUC_{training}$ = 0.830, AUC_{test} = 0.779, and $AUC_{validation}$ = 0.711), radiomics ($AUC_{training}$ = 0.810, AUC_{test} = 0.710, and $AUC_{validation}$ = 0.839), and clinical ($AUC_{training}$ = 0.780, AUC_{test} = 0.685, and $AUC_{validation}$ = 0.695) models, the combined model based on the GP algorithm, which consisted of four DL features, five radiomics features, and two clinical variables, not only demonstrated the highest diagnostic efficacy ($AUC_{training}$ = 0.949, AUC_{test} = 0.877, and $AUC_{validation}$ = 0.914) but also led to an improvement in risk reclassification of the TP53 mutation ($NIR_{training}$ = 66.38%, 56.98%, and 83.48%, NIR_{test} = 50.72%, 80.43%, and 89.49%, and $NIR_{validation}$ = 64.58%, 87.50%, and 120.83%, respectively). In addition, the combined model exhibited good agreement and clinical utility in calibration curves and DCA analyses, respectively.

Conclusions A prediction model based on the GP algorithm and consisting of DL and radiomics features of DWI as well as clinical variables can effectively assess TP53 mutation in EC.

Keywords Endometrial cancer, TP53 mutation, Diffusion-weighted imaging, Radiomics, Deep learning

[†]Lei Shen and Bo Dai contributed equally to this work.

*Correspondence:
Yaping Wu
ypwu@ha.edu.cn
[1]Department of Radiology, Henan Provincial People's Hospital & Zhengzhou University People's Hospital, Zhengzhou, Henan, China

Introduction

The TP53 gene is thought to be the most commonly mutated gene in human cancers [1]. The occurrence and development of endometrial cancer, one of the most lethal malignant tumors of the female reproductive system, are also closely related to the mutation status of the TP53 gene [2, 3]. Previous studies have shown that a TP53 mutation frequently imply greater aggressiveness, drug resistance, and poorer prognosis in EC patients [4, 5]. Immunohistochemistry (IHC) is a relatively common approach for assessing TP53 mutations in clinical practice [6]. However, this method is not only potentially costly to the patient but also often only performed when adequate sample sizes are obtained postoperatively, which limits its prevalence. Consequently, developing a cost-effective and noninvasive method to detect mutations in the TP53 gene may benefit patients with EC.

Diffusion weighted imaging (DWI) is a quantitative magnetic resonance imaging (MRI) technique that can not only effectively reveal the location and number of lesions but also quantify the cellular density of tissues using apparent diffusion coefficient (ADC) values [7]. Previous works have demonstrated that DWI can play an important role in the assessment of various EC histological features, including lymph node metastasis, risk stratification, and microsatellite instability [8–10]. Radiomics, a significant breakthrough in the evolution of information technology, enables the discovery and utilization of higher-order image features that may be difficult for radiologists to identify by eye; these imaging features in turn are effective markers for the diagnosis, evaluation and prognostic prediction of relevant diseases [11]. There have been several investigations proving the value of radiomics in the clinical classification [12], identification of lymphovascular spatial invasion (LVSI) [13], and prediction of recurrence [14] of EC patients. Deep learning (DL) is a deeper evolution of information technology that extracts image information with multilevel convolutional neural networks (CNNs), yielding highly robust and selective features that enhance the diagnostic and assessment reliability of diseases [15]. Several studies have illustrated that DL is effective in terms of lesion identification [16], myometrial infiltration depth assessment [17], and staging in EC patients [18]. To date, the predictive value of MRI-based radiomic models for TP53 expression status has been demonstrated in rectal and breast cancer patients [19, 20]. For EC patients, only Jiang et al. has explored the role of texture analysis in TP53 status assessment [21]. To the best of our knowledge, there are no studies of deep learning combined with radiomics differences for TP53 mutant and TP53 wild-type EC patients.

Therefore, the aim of this study was to investigate the feasibility of a predictive model based on DWI features extracted from DL and radiomics in combination with clinical variables to discriminate between TP53 mutant and TP53 wild-type EC patients, with the aim of introducing a novel and non-invasive pretreatment tool into clinical practice.

Materials and methods

Study participants

This retrospective study was conducted with the approval of the local ethics committee (No.2021150), and informed consent was waived for all participants. The initial dataset included 225 patients from the centre I and 80 patients from centre II who underwent pelvic MRI for suspected EC between January 2020 and September 2023. Patients were excluded based on the following criteria: (1) pathologically confirmed non-EC ($n = 25$); (2) missing TP53 status assessment ($n = 70$); (3) prior radiotherapy, chemotherapy, or surgery before MRI examination ($n = 16$); (4) absence of DWI sequences or those of insufficient quality for analysis ($n = 18$); and (5) incomplete clinical or histopathological information ($n = 21$). Ultimately, 115 patients from centre I and 40 patients from centre II were included in the study. The data from Centre I were sorted in chronological order, with the first 70% of patients in the TP53-mutant and TP53-wild groups designated as the training set and the last 30% of patients in both groups used as the test set. The data from Centre II were designated as the external validation cohort.

At baseline, clinical variables, including ADC value, age, maximum lesion diameter, levels of carcinoembryonic antigen (CEA) and carbohydrate antigen 153 (CA 153), and International Federation of Gynecology and Obstetrics (FIGO) stage (Vison 2009), were recorded.

Scanning process

All scans were conducted using either a 1.5 T (Optima MR360, GE, Milwaukee, USA) or 3.0 T (Signa HDxt/Discovery 750 W, GE, Milwaukee, USA) MRI scanner. All patients were positioned supine, feet first, and scanned from the symphysis pubis to the anterior superior iliac spine. Axial T2-weighted imaging and DWI (b = 800 s/mm^2) data were obtained for subsequent analysis. For more detailed information, please refer to Table 1.

Image pre-processing

To minimize the impact of using various MRI scanners, the data from the DWI scans were initially preprocessed and enhanced, which included transformation of Digital Imaging and Communications in Medicine (DICOM) files into the Neuroimaging Informatics Technology Initiative (NIfTI) file format, alignment, resampling to achieve uniform isotropic resolution (1 mm^3), and random axis mirror flipping.

Table 1 MRI acquisition parameters

Scanner	GE Healthcare Optima MR360 1.5T	GE Healthcare Signa HDxt 3.0T	GE Healthcare Discovery 750 W 3.0T
DWI	Axial, 2D EPI	Axial, 2D EPI	Axial, 2D EPI
TR/TE (ms)	2360/76.4	3327/72.4	4120/68.9
Flip angle (°)	90	90	90
Slice/Gap (mm)	5/1	5/1	5/1
Matrix	280×208	260×280	128×128
FOV (mm)	380	380	380
b (s/mm^2)	800	800	800
T2WI	Axial, 2D FSE	Axial, 2D FSE	Axial, 2D FSE
TR/TE (ms)	4588/80.2	3748/75.8	4378/88.6
Flip angle (°)	120	112	110
Slice/Gap (mm)	5/1	5/1	5/1
Matrix	320×256	382×479	384×384
FOV (mm)	380	380	380

DWI = diffusion-weighted imaging, T2WI = T2-weighted imaging, TR = repetition time, TE = echo time, EPI = echo-planar imaging, FSE = fast spin echo, FOV = feld of view

Tumor segmentation

Tumor segmentation was performed using ITK-SNAP software (version 3.8.0; http://www.itksnap.org). Initially, a radiologist with 6 years of experience (reader 1) manually delineated regions of interest (ROIs) encompassing the tumors, slice-by-slice, on the axial DWI images. Following this, the software's three-dimensional functionality was utilized to integrate the ROIs from all the slices of each tumor, resulting in the creation of 3D volumes of interest (VOIs). The image processing workflow is presented in Fig. 1.

ADC calculation

The ADC values of the tumor VOI were calculated using the equation:

$$S_b/S_0 = \exp(-b \times ADC)$$

where S_b and S_0 represent the signal intensities at b values of 0 s/mm^2 and 'b', respectively, in which 'b' represents the diffusion sensitizing factor.

Radiomic feature extraction

The standardized radiomics analysis workflow software package PyRadiomics (version 2.1.2; https://pyradiomics.readthedocs.io) was employed for radiomic feature extraction [22]. First, all images were normalized using z score normalization with a fixed bin width of 20 and voxel size resampling to 1 mm^3 [23]. Then, 12 filters, including Original, Laplacian Sharpening, wavelet, AdditiveGaussianNoise, Normalize, BoxMean, BinomiaIBIurlmage, ShotNoise, LoG, Boxsigmalmage, CurvatureFlow, and DiscreteGaussian, were applied. As a result, 2264 radiomic features were extracted from each of the VOIs.

DL feature extraction

We designed a DL feature extraction model based on multiscale CNNs to refine and extract features from tumors of different grades. The model consists of 8 convolutional layers with ReLU activation (red), 3 maximal pooling layers (blue), 3 upsampling layers (yellow), and a full sampling layer (purple). The multiscale network design consists of five bottom-up layers (convolution 1-convolution 5) and three merged layers (convolution 6-convolution 8), where the larger scale focuses on the large target of the tumor and the smaller scale obtains richer contextual information. During each round of training of the network, the DWI images are preprocessed separately and fed into the model, and DL features are output through the fully connected layers [24]. The above process is shown in Fig. 2. A self-written code based on Python (version 3.1.0) was used for all these operations, and 128 DL features were finally extracted from each DWI image (Supplementary 1).

Feature selection

All feature extraction is carried out in the training set, and then agreement in the inter- and intraobserver characteristics was assessed using the DWI images of 20 randomly selected patients in training set. Two weeks later, intraobserver reproducibility was estimated by having reader 1 repeat the segmentation of the VOIs from the DWI scans. To assess interobserver reproducibility, Reader 2 (another radiologist with 17 years of experience) also independently delineated the VOIs and extracted features in the same way. For the radiomics and DL features, robust features were initially selected if their interobserver and intraobserver interclass correlation coefficients (ICCs) > 0.75. All features were then normalized using the z score method. The Mann–Whitney U test ($\alpha = 0.05$), Select K Best algorithm (K = 30), and Least

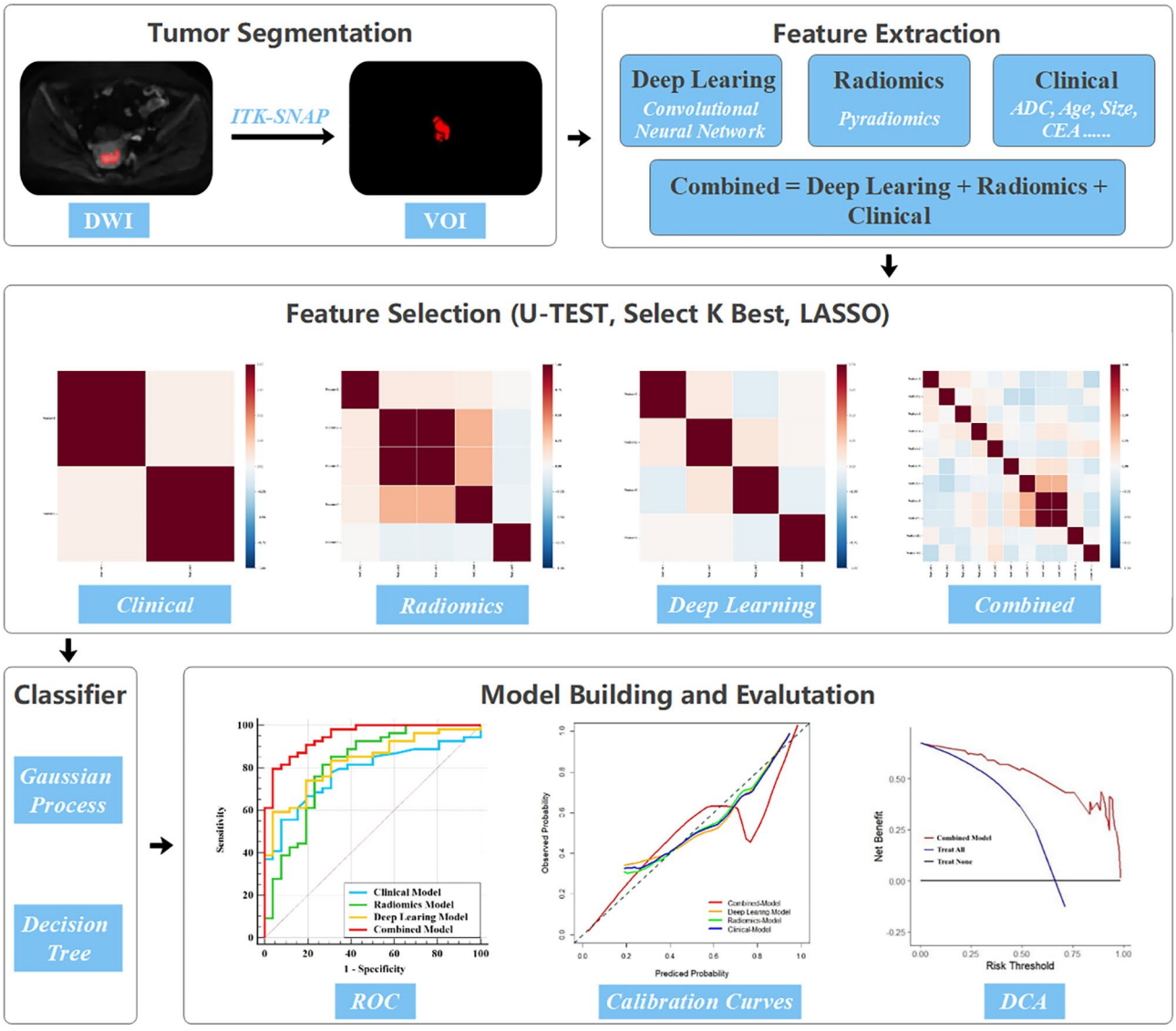

Fig. 1 Image processing workflow

Absolute Shrinkage and Selection Operator (LASSO) algorithm (α = 0.01, max_iter = 1000, and n-folds = 5) were then used to eliminate redundant features.

Model development

Because of their simplicity, effectiveness and popularity, two algorithms, Gaussian Process (GP) and Decision Tree (DT), were used to develop the model. For the GP algorithms, the following parameters were set: Length scale = 1.0, Optimizer restart count = 10, Threshold = 0.5; For the DT algorithms, the parameter values were as follows: Category weight = None, Indicator = Gini, Maximum depth = 4, Minimum number of leaf node samples = 1, Minimum cut sample room = 2, Cutting method = Best, Classification threshold = 0.5. Eight (2 × 4) prediction models were generated, including 3 independent models (based on clinical characteristics and radiomics and DL features alone) and 1 combined model (clinical + radiomics + DL) for the two algorithms. The final predictive model was determined by meeting certain criteria to avoid overfitting and ensure robustness; specifically, the area under the curve (AUC) of the models in the training and test sets should be maximized while ensuring that the difference between the two sets is less than 0.1 [25].

TP53 status assessment

All lesion specimens were sent to the pathology center of our institution for assessment of TP53 status by IHC staining. The presence of brown granules in the nucleus was used to assess the expression of TP53 [6]. This work was carried out independently by two experienced

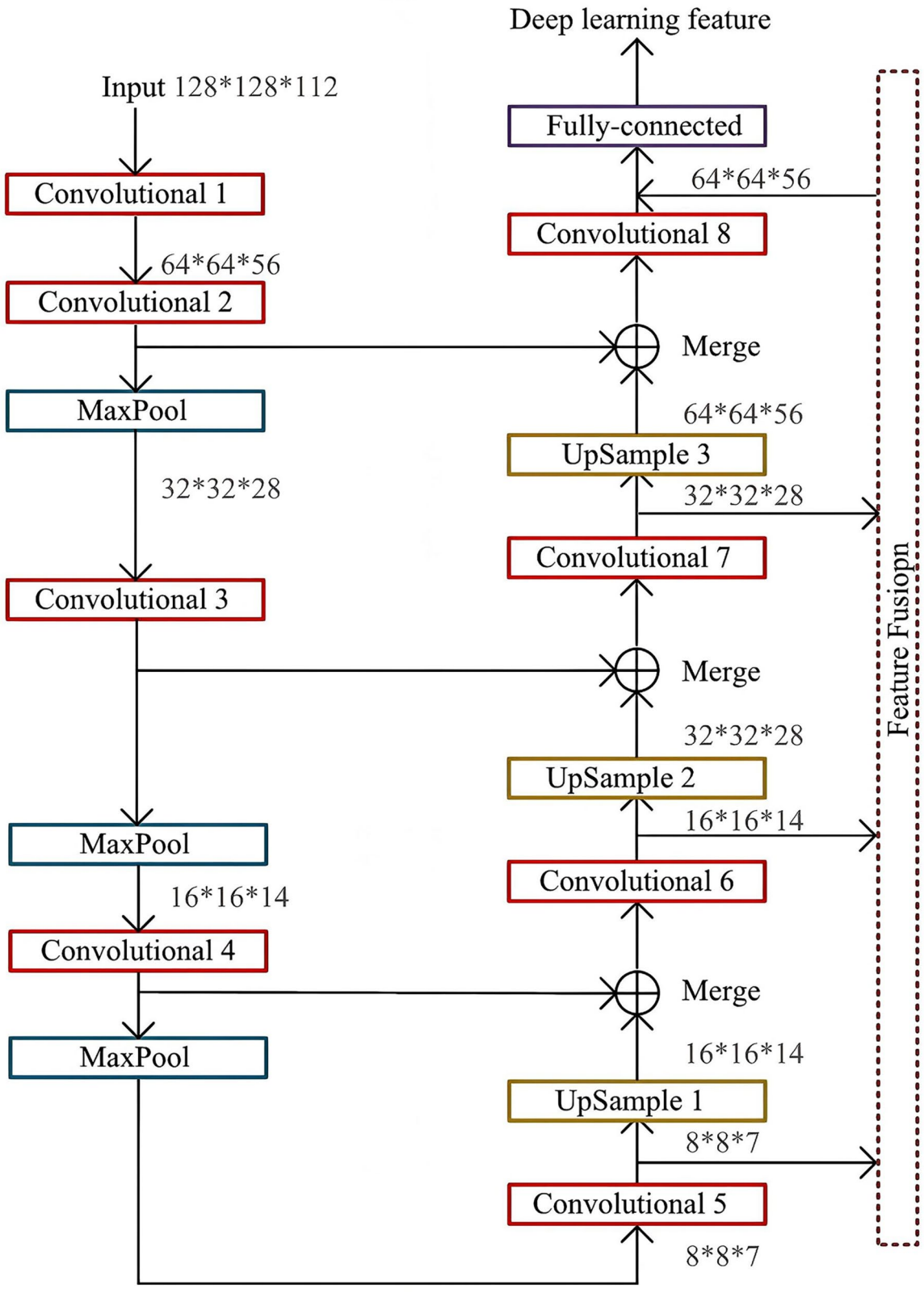

Fig. 2 Schematic of deep learning feature generation. Red represents the eight convolutional layers with ReLU activation, blue represents the three maximal pooling layers, yellow represents the three upsampling layers, and purple represents the one full sampling layer

pathologists, and disagreements, if any, were resolved by discussion.

Statistical analysis

Data analyses were performed using SPSS 26.0 (IBM) and RStudio 4.3.1 (R Foundation) software. Statistical significance was established if $P < 0.05$. Differences in variables between groups were analyzed using the Mann–Whitney U test, independent samples t test, or chi-square test. The AUC, DeLong analysis, and net reclassification index (NRI) were used to evaluate the diagnostic performance of the models. Calibration curves and decision curve analysis (DCA) were employed for model validation and net clinical benefit assessment, respectively.

Results

Clinical variables

The training set comprised 80 patients, including 54 mutation and 26 wild cases. The test set comprised 35 patients, including 23 mutation and 12 wild cases. The external validation set comprised 40 patients, including 24 mutation and 16 wild cases. None of the clinical variables, including ADC value, age, maximum diameter, CEA, CA153, and FIGO stage, were significantly different in the training, test, and external validation sets (*P*-value all > 0.05). The ADC values in the TP53-mutant group were significantly lower than those in the TP53-wild-type group, a trend observed consistently in the training, test, and external validation sets. The details of the clinical variables are summarized in Table 2.

Prediction model

The final clinical model was a GP-based model comprising two clinical variables, ADC value and FIGO stage. The final radiomic model was a GP-based model comprising five radiomic features, wavelet_glrlm_wavelet-LHH-RunVariance, wavelet_glcm_wavelet-LLH-Idmn, wavelet_glcm_wavelet-LLH-Idn, log_glcm_log-sigma-4-0-mm-3D-Correlation, and wavelet_glszm_wavelet-HHL-GrayLevelVariance. The final DL model was a

Table 2 Summary of characteristics in differents sets

Training set	Mutation	Wild	*P* value	*P**value
Number	*n* = 54	*n* = 26		
ADC value	1.02 ± 0.39	1.24 ± 0.27	0.004	0.448
Age (year)	56.37 ± 7.99	57.46 ± 8.90	0.598	0.187
Maximum diameter (mm)	31.85 (9.83, 52.55)	28.95 (7.48, 40.30)	0.337	0.686
CA-153 (u/ml)	8.66 (7.21, 10.46)	9.07 (7.46, 14.07)	0.374	0.363
CEA (ng/ml)	1.65 (1.33, 2.29)	1.53 (1.29, 1.83)	0.182	0.248
FIGO stage			0.055	0.593
Early (≤ II)	47 (87.04%)	26 (100.00%)		
Advanced (> II)	7 (12.96%)	0 (0.00%)		
Test set				
Number	*n* = 23	*n* = 12		
ADC value	0.90 ± 0.20	1.18 ± 0.44	0.016	0.448
Age (year)	54.26 ± 8.44	52.75 ± 6.02	0.546	0.187
Maximum diameter (mm)	41.40 (10.00, 48.90)	18.60 (8.35, 28.85)	0.085	0.686
CA-153 (u/ml)	8.48 (4.60, 9.69)	9.79 (8.17, 11.56)	0.097	0.363
CEA (ng/ml)	1.37 (0.68, 2.12)	1.72 (1.43, 2.30)	0.208	0.248
FIGO stage			0.630	0.593
Early (≤ II)	22 (95.65%)	11 (91.67%)		
Advanced (> II)	1 (4.35%)	1 (8.33%)		
Validation set				
Number	*n* = 24	*n* = 16		
ADC value	1.00 ± 0.35	1.19 ± 0.39	0.117	0.448
Age (year)	56.33 ± 6.89	55.06 ± 9.07	0.618	0.187
Maximum diameter (mm)	42.35 (12.65, 56.58)	24.25 (8.05, 47.10)	0.252	0.686
CA-153 (u/ml)	8.11 (7.05, 9.24)	9.74 (7.53, 16.49)	0.044	0.363
CEA (ng/ml)	1.72 (0.89, 2.57)	1.56 (1.16, 1.87)	0.720	0.248
FIGO stage			0.051	0.593
Early (≤ II)	19 (79.17%)	16 (100.00%)		
Advanced (> II)	5 (20.83%)	0 (0.00%)		

FIGO = International Federation of Gynecology and Obstetrics, LNM = lymph node metastasis, CEA = carcinoembryonic antigen, CA-125 = carbohydrate antigen 125, and CA-153 = carbohydrate antigen 153. P = the result of univariable analyses between mutation and wild; P* = the results of the univariable analyses between each clinical parameter in different sets

GP-based model comprising four DL features, the 56th, 59th, 76th, and 78th. The final combined model (clinical + radiomics + DL) was a GP-based model comprising the two clinical variables, five radiomic features, and four DL features listed above (Fig. 3).

Receiver operating characteristic (ROC) curve and NRI analysis

Compared with the clinical, radiomics, and DL, models, the combined model (clinical + radiomics + DL) exhibited improved diagnostic performance in the training, test, and external validation sets, with AUC values of 0.949,

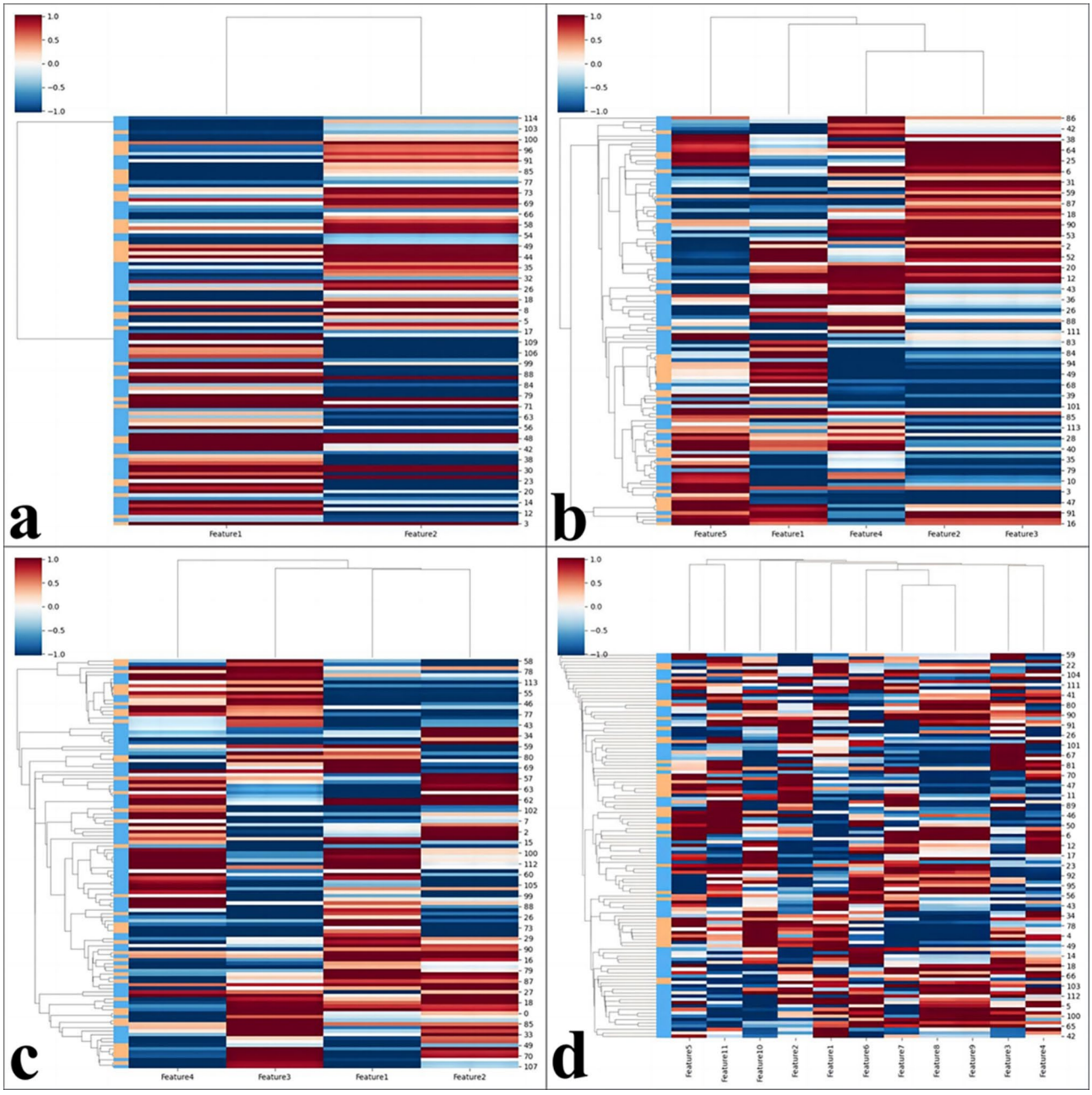

Fig. 3 Feature heatmap for different models. Figure **a** is the heatmap of features in the clinical model, where features 1–2 are FIGO stage and ADC, respectively. Figure **b** is the heatmap of features in the radiomic model, where features 1–5 are log_glcm_log-sigma-4-0-mm-3D-Correlation, wavelet_glcm_wavelet-LLH-Id, wavelet_glcm_wavelet-LLH-Idm, wavelet_glrlm_wavelet-LHH-Run Variance, and wavelet_glszm_ wavelet-HHL-GrayLevel Variance, respectively. Figure **c** is the heatmap of features in the DL model, where features 1–4 are the 56th, 59th, 76th, and 78th features, respectively. Figure **d** is the heatmap of features in the combined model, where features 1–11 are FIGO stage, ADC, 56th, 59th, 76th, 78th, wavelet_glrlm_wavelet-LHH-RunVariance, wavelet_glcm_wavelet-LLH-Idmn, wavelet_glcm_wavelet-LLH-Idn, log_glcm_log-sigma-4-0-mm-3D-Correlation, and wavelet_ glszm_wavelet-HHL-GrayLevelVariance, respectively

0.877, and 0.914, respectively. In the training set, the differences in the AUCs between the combined model and the clinical (AUC = 0.780), radiomics (AUC = 0.810), and DL (AUC = 0.830) models were all significant (Z = 3.658, 2.956, and 2.720, $P < 0.001$, = 0.003, and = 0.007, respectively). In the test set, a significant difference in AUC was observed between the combined model and the radiomic (AUC = 0.710) model (Z = 2.185, $P = 0.029$), whereas no significant differences were observed in the AUC between the combined model and the clinical (AUC = 0.685) and DL (AUC = 0.779) models (Z = 1.614, 1.016, $P = 0.106$, 0.310, respectively). In the external validation set, the difference in the AUC between the combined model and the clinical (AUC = 0.695) and DL (AUC = 0.711) models were all significant (Z = 2.813, 2.438, $P = 0.005$, 0.015, respectively), whereas no significant differences were observed in the AUC between the combined model and the radiomics (AUC = 0.839) model (Z = 1.307, $P = 0.191$, respectively, Table 3; Fig. 4).

Compared with the clinical, radiomics, and DL, models, the combined model (clinical + radiomics + DL) demonstrated superior risk reclassification of the TP53 status. In the training set, the NRIs were 83.48% (95% CI: 55.74 ~ 111.21), 56.98% (95% CI: 29.83 ~ 84.13), and 66.38% (95% CI: 38.62 ~ 94.14), respectively (P all < 0.001). In the test set, the NRIs were 89.49% (95% CI: 47.92 ~ 131.07), 80.43% (95% CI: 44.39 ~ 116.48), and 50.72% (95% CI: 8.59 ~ 92.86), respectively ($P < 0.001$, < 0.001, and = 0.018, respectively). In the external validation set, the NRIs were 120.83% (95% CI: 70.64 ~ 171.02), 87.50% (95% CI: 30.37 ~ 144.63), and 64.58% (95% CI: 26.88 ~ 102.29), respectively ($P < 0.001$, = 0.003, and < 0.001, respectively).

Model validation and clinical benefit

In the training, test, and external validation sets, the calibration curves revealed that the combined model (clinical + radiomics + DL) exhibited good agreement between the predicted values and actual observations, while DCA showed that the combined model (clinical + radiomics + DL) offered a reliable clinical benefit in assessing the TP53 status of EC patients (Figs. 5 and 6).

Table 3 Performance of prediction models

Model	Classifier	AUC (95% CI)	Sensitivity	Specificity	Youden
Training set					
Clinical	**GP**	0.780 (0.673–0.865)	55.56%	92.31%	47.86%
	DT	0.902 (0.815–0.957)	85.19%	80.77%	65.95%
Radiomics	**GP**	0.810 (0.707–0.889)	81.48%	73.08%	54.56%
	DT	0.949 (0.876–0.986)	98.15%	80.77%	78.92%
DL	**GP**	0.830 (0.729–0.905)	59.26%	96.15%	55.41%
	DT	0.962 (0.893–0.992)	88.89%	96.15%	85.04%
Combined	**GP**	0.949 (0.875–0.986)	79.63%	96.15%	75.78%
	DT	0.990 (0.937–1.000)	94.44%	96.15%	90.60%
Test set					
Clinical	**GP**	0.685 (0.506–0.831)	95.65%	50.00%	45.65%
	DT	0.522 (0.347–0.693)	34.78%	75.00%	9.78%
Radiomics	**GP**	0.710 (0.533–0.850)	82.61%	66.67%	49.28%
	DT	0.737 (0.561–0.871)	95.65%	41.67%	37.32%
DL	**GP**	0.779 (0.607–0.901)	56.52%	91.67%	48.19%
	DT	0.634 (0.455–0.790)	60.87%	75.00%	35.87%
Combined	**GP**	0.877 (0.722–0.963)	73.91%	91.67%	65.58%
	DT	0.614 (0.435–0.773)	78.26%	50.00%	28.26%
Validation set					
Clinical	**GP**	0.695 (0.530–0.831)	83.33%	56.25%	39.58%
	DT	0.698 (0.533–0.833)	50.00%	81.25%	31.25%
Radiomics	**GP**	0.839 (0.688–0.936)	83.33%	81.25%	64.58%
	DT	0.888 (0.748–0.966)	87.50%	81.25%	68.75%
DL	**GP**	0.711 (0.546–0.843)	54.17%	87.50%	41.67%
	DT	0.849 (0.700–0.942)	83.33%	87.50%	70.83%
Combined	**GP**	0.914 (0.782–0.979)	70.83%	100.00%	70.83%
	DT	0.829 (0.677–0.930)	79.17%	81.25%	60.42%

CI = confidence interval, DL = deep learning, GP = gaussian process, DT = decision tree, combined = clinical + DL + Radiomics. The bold typeface indicates the final model

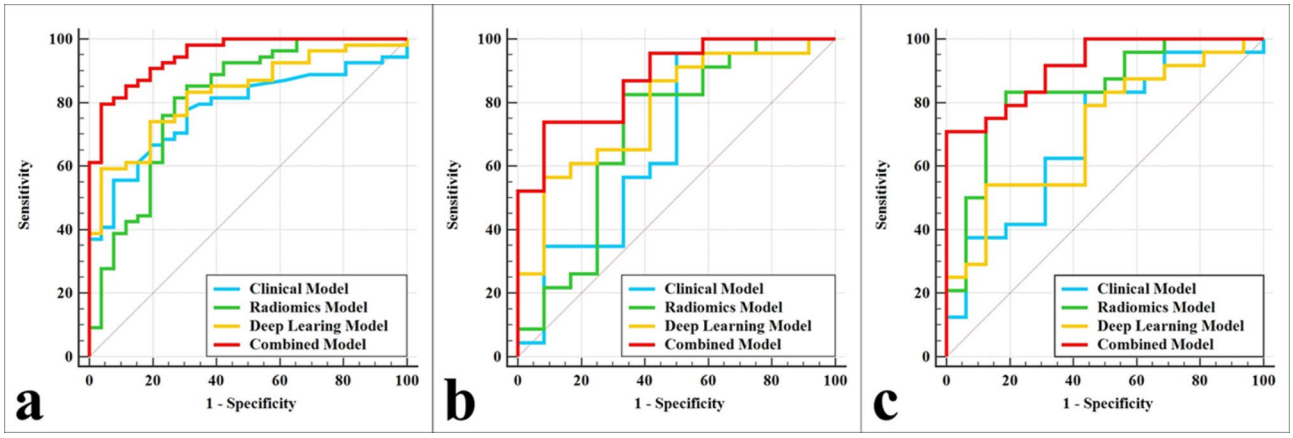

Fig. 4 Areas under the receiver operating characteristic curves of the clinical (blue), deep learning (DL) (yellow), radiomic (green), and combined (clinical + DL + radiomic + ADC, red) models in the training (**a**) and test (**b**) sets

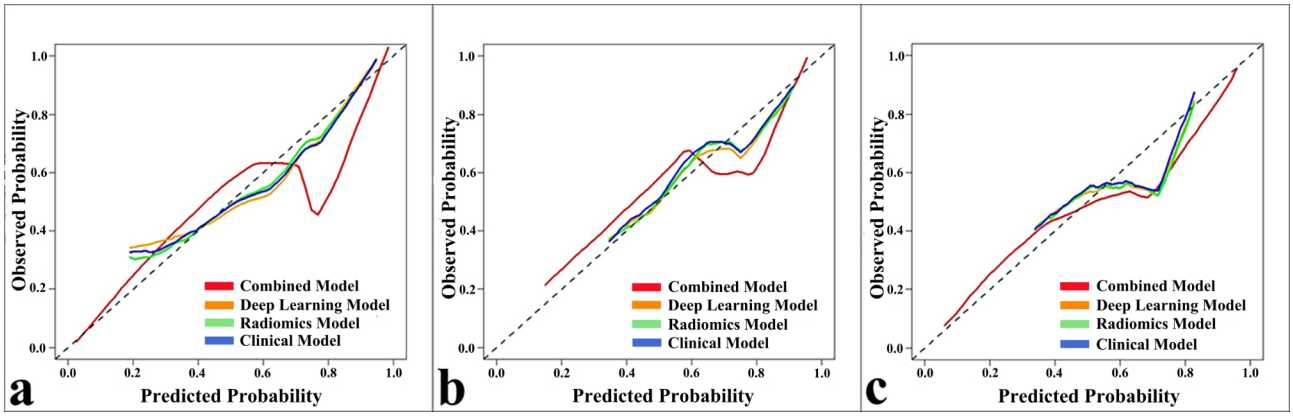

Fig. 5 Calibration curves of the clinical (blue), deep learning (DL) (yellow), radiomic (green), and combined (clinical + DL + radiomic, red) models in the training (**a**) and test (**b**) sets

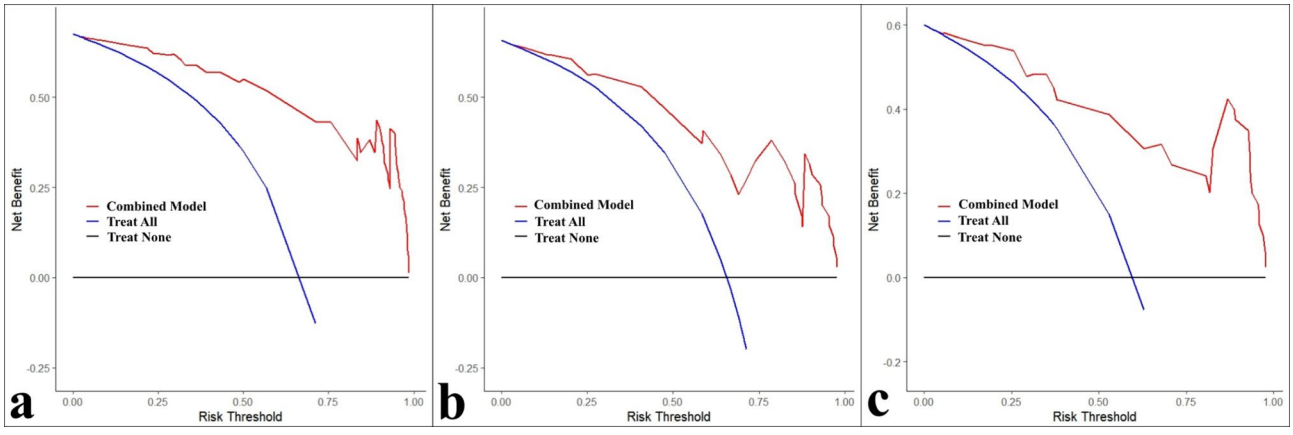

Fig. 6 Decision curve analysis results for the combined (clinical + DL + radiomic) model in the training (**a**) and test (**b**) set

Discussion

The results of this study showed that the clinical model, DWI-based radiomic model, and DWI-based DL model were all capable of assessing the TP53 mutation status in EC patients. In addition, a combined model consisting of two clinical variables, five radiomic features and four DL features demonstrated improved diagnostic performance and risk reclassification over the single models described above.

During the construction of the clinical model, a total of seven clinical variables (ADC value, age, maximum diameter, CEA level, CA 153 level, FIGO stage and histological grade) were included. However, after screening, only FIGO stage and ADC value were found to have a robust relationship with the TP53 mutation status of EC patients. The FIGO stage is mainly used to reflect the degree of invasion of the lesion into the surrounding tissues; usually, the more malignant the lesion is and the more severe the invasion of the surrounding tissues is, the higher the FIGO stage is [26]. The ADC is a quantitative parameter derived from DWI to reflect the rapidity of the diffusion of water molecules moving within the tissue. Previous studies have shown that the more malignant the lesion is, the more significant the restriction of water molecule diffusion within it, and the lower the ADC value [27]. An increase in the expression of mutant TP53 may lead to higher proliferative activity, poorer differentiation ability, and higher malignancy of EC cells [2, 4]. These factors may exacerbate the degree of invasion of EC peripheral tissues and the degree of internal water molecule diffusion restriction to varying degrees, resulting in a robust correlation between FIGO stage and ADC values and TP53 expression in EC patients. Furthermore, a clinical model was constructed based on the above clinical variables, but the results showed that although the model was able to assess the TP53 mutation status in EC patients, the diagnostic efficacy was low. This is consistent with previous studies, and the reason for this result may be related to the limited information and low specificity for TP53 from single clinical variables [19, 21].

In this research, we used DL and radiomics to extract features of DWI images and built corresponding models. We found that both the DL and radiomic models effectively assessed the TP53 mutation status of EC patients, and the diagnostic efficacy of the DL model was higher than that of the radiomic model in both the training and test sets. This result is consistent with previous studies [28, 29], demonstrating that DL is feasible and relatively superior to radiomics in discriminating between TP53-mutant and TP53-wild-type EC patients. One possible reason for the superior performance of DL over radiomics is that compared to radiomics, which extracts a relatively fixed amount of feature information, CNNs have a more complex and scalable structure, which makes them capable of not only extracting deeper and more subtle features from an image but also perhaps mimicking the elusive subconscious process that occurs when radiologists interpret DWI images [30, 31]. As a result, DL is able to efficiently detect subtle differences between TP53-mutant and TP53-wild-type EC, leading to improved diagnostic efficacy. In addition, we also constructed a combined model consisting of clinical variables, DL features, and radiomic features and found that it showed improved diagnostic efficacy and reclassification ability to varying degrees over the clinical, radiomic, and DL models. This result is consistent with previous work showing that a combined model that integrates multidimensional validation information can offer a more comprehensive understanding of the lesion than a single-modality model [32, 33], thereby enabling a more accurate distinction between TP53-mutant and TP53-wild-type EC.

The suitability of the algorithm has a very important impact on the diagnostic performance of the prediction model. The DT algorithm is a supervised learning algorithm that uses the idea of classification to construct a mathematical model based on the characteristics of the data to screen data and make decisions [34]. The GP algorithm is an unsupervised learning algorithm that mainly through the use of a measure of homogeneity between data points as a kernel function, constructs a covariance function through the training samples to obtain the joint probability density and then find the new data prediction value [35]. In this study, due to the simplicity, effectiveness and popularity of the above two algorithms, we chose them to construct distinct models. The results showed that both in the single models (individual clinical, radiomics, and DL models) and in the combined model (clinical + radiomics + DL), the GP algorithm demonstrated better stability and reliability than the DT algorithm, achieving higher diagnostic efficacy in both the training and test sets, but the difference in diagnostic performance between the two was small. This is similar to previous findings, and we speculate that this may be related to the fact that GP algorithms are more suitable for solving small-sample, nonlinear, and noisy problems [36]. In addition, considering that different algorithms have their own optimal applicability scenarios and different diseases have their own unique pathophysiological characteristics, future studies may wish to consider applying as many algorithms as possible that can be used for modeling when conditions allow.

This study has certain limitations. First, although the data in the present study came from two centres and a training set, a test set and an external validation set were set up, the sample size is still small, which may affect the clinical extension of the prediction model to some extent. Second, the DWI scanning protocols used in this study were implemented in different scanners, which may have negatively affected the results despite standardized preprocessing. Third, we only implemented the DT and GP algorithms, which may have limited our ability to build a predictive model with better diagnostic performance in this study. In the future, we will further expand the sample size, conduct prospective multicentre studies, optimise the imaging solution and integrate more algorithms, with a view to obtaining more accurate and reliable

results, which will lead to the promotion and application of this prediction model in clinical practice.

Conclusion

In conclusion, our study revealed that a prediction model based on the GP algorithm and consisting of DL and radiomics DWI features as well as clinical variables can effectively assess TP53 mutation status in EC. With the continuous improvement of the relevant algorithms and further in-depth studies, the model is expected to provide a reliable non-invasive assessment of TP53 mutations in EC patients, which in turn will facilitate the development of intelligent clinical practice for relevant patients.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13424-5.

Supplementary Material 1

Acknowledgements
We acknowledge the support received from the Henan Province Medical Science and Technology Project.

Author contributions
YP. W and L.S wrote the main manuscript text; TY.Y, FS.Y, and SW. D collected data; and B. D prepared Figs. 1, 2, 3, 4, 5 and 6. All authors reviewed the manuscript.

Funding
The Henan Provincial Science and Technology Research Projects (222102310627).

Data availability
The datasets during and/or analysed during the current study available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This retrospective study was conducted with the approval of the Henan Provincial People's Hospital ethics committee (No.2021150), and informed consent was waived for all participants.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 7 February 2024 / Accepted: 1 January 2025
Published online: 09 January 2025

References

1. Voskarides K, Giannopoulou N. The role of TP53 in adaptation and evolution. Cells. 2023;12(3):512. https://doi.org/10.3390/cells12030512.
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7–33. https://doi.org/10.3322/caac.21708.
3. Bell DW, Ellenson LH. Molecular Genetics of Endometrial Carcinoma. Annu Rev Pathol. 2019;14:339–67. https://doi.org/10.1146/annurev-pathol-020117-043609.
4. Coll-de la Rubia E, Martinez-Garcia E, Dittmar G, Gil-Moreno A, Cabrera S, Colas E. Prognostic biomarkers in Endometrial Cancer: a systematic review and Meta-analysis. J Clin Med. 2020;9(6):1900. https://doi.org/10.3390/jcm9061900.
5. hiel KW, Devor EJ, Filiaci VL, et al. TP53 sequencing and p53 immunohistochemistry predict outcomes when Bevacizumab is added to Frontline Chemotherapy in Endometrial Cancer: an NRG Oncology/Gynecologic Oncology Group Study. J Clin Oncol. 2022;40(28):3289–300. https://doi.org/10.1200/JCO.21.02506.
6. Singh N, Piskorz AM, Bosse T, et al. p53 immunohistochemistry is an accurate surrogate for TP53 mutational analysis in endometrial carcinoma biopsies. J Pathol. 2020;250(3):336–45. https://doi.org/10.1002/path.5375.
7. Fujii S, Gonda T, Yunaga H. Clinical utility of diffusion-weighted imaging in gynecological imaging: revisited. Invest Radiol. 202310.1097/RLI.0000000000001004.
8. Arian A, Easa AM, Arab-Ahmadi M. Diagnostic value of diffusion-weighted magnetic resonance imaging in discriminating between metastatic and non-metastatic pelvic lymph nodes in endometrial cancer. Acta Radiol. 2020;61(11):1580–6. https://doi.org/10.1177/0284185120906660.
9. Meng N, Fang T, Feng P, et al. Amide Proton transfer-weighted imaging and multiple models diffusion-weighted imaging facilitates preoperative risk stratification of early-stage endometrial carcinoma. J Magn Reson Imaging. 2021;54(4):1200–11. https://doi.org/10.1002/jmri.27684.
10. Ma C, Tian S, Song Q, et al. Amide Proton transfer-weighted imaging combined with Intravoxel Incoherent Motion for evaluating microsatellite instability in Endometrial Cancer. J Magn Reson Imaging. 2023;57(2):493–505. https://doi.org/10.1002/jmri.28287.
11. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J Nucl Med. 2020;61(4):488–95. https://doi.org/10.2967/jnumed.118.222893.
12. Liu XF, Yan BC, Li Y, Ma FH, Qiang JW. Radiomics nomogram in aiding preoperatively dilatation and curettage in differentiating type II and type I endometrial cancer. Clin Radiol. 2023;78(2):e29–36. https://doi.org/10.1016/j.crad.2022.08.139.
13. Long L, Sun J, Jiang L, et al. MRI-based traditional radiomics and computer-vision nomogram for predicting lymphovascular space invasion in endometrial carcinoma. Diagn Interv Imaging. 2021;102(7–8):455–62. https://doi.org/10.1016/j.diii.2021.02.008.
14. Zhang K, Zhang Y, Fang X, Dong J, Qian L. MRI-based radiomics and ADC values are related to recurrence of endometrial carcinoma: a preliminary analysis. BMC Cancer. 2021;21(1):1266. https://doi.org/10.1186/s12885-021-08988-x.
15. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in Medical Image Analysis. Adv Exp Med Biol. 2020;1213:3–21. https://doi.org/10.1007/978-3-030-33128-3_1.
16. Urushibara A, Saida T, Mori K, et al. The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: a comparison with radiologists. BMC Med Imaging. 2022;22(1):80. https://doi.org/10.1186/s12880-022-00808-3.
17. Chen X, Wang Y, Shen M, et al. Deep learning for the determination of myometrial invasion depth and automatic lesion identification in endometrial cancer MR imaging: a preliminary study in a single institution. Eur Radiol. 2020;30(9):4985–94. https://doi.org/10.1007/s00330-020-06870-1.
18. Mao W, Chen C, Gao H, Xiong L, Lin Y. A deep learning-based automatic staging method for early endometrial cancer on MRI images. Front Physiol. 2022;13:974245. https://doi.org/10.3389/fphys.2022.974245.
19. Sun K, Zhu H, Chai W, Yan F. TP53 mutation estimation based on MRI Radiomics analysis for breast Cancer. J Magn Reson Imaging. 2023;57(4):1095–103. https://doi.org/10.1002/jmri.28323.
20. Zhong X, Peng J, Shu Z, Song Q, Li D. Prediction of p53 mutation status in rectal cancer patients based on magnetic resonance imaging-based nomogram: a study of machine learning. Cancer Imaging. 2023;23(1):88. https://doi.org/10.1186/s40644-023-00607-1. Published 2023 Sep 18.
21. Jiang X, Jia H, Zhang Z, Wei C, Wang C, Dong J. The feasibility of combining ADC value with texture analysis of T2WI, DWI and CE-T1WI to Preoperatively predict the expression levels of Ki-67 and p53 of endometrial carcinoma. Front Oncol. 2022;11:805545. https://doi.org/10.3389/fonc.2021.805545.
22. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic phenotype. Cancer Res. 2017;77(21):e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339.
23. Shin J, Seo N, Baek SE, et al. MRI Radiomics Model predicts pathologic complete response of rectal Cancer following Chemoradiotherapy. Radiology. 2022;303(2):351–8. https://doi.org/10.1148/radiol.211986.

24. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234–241. https://doi.org/10.1007/978-3-319-24574-428
25. Zhou Y, Ma XL, Zhang T, Wang J, Zhang T, Tian R. Use of radiomics based on 18F-FDG PET/CT and machine learning methods to aid clinical decision-making in the classification of solitary pulmonary lesions: an innovative approach. Eur J Nucl Med Mol Imaging. 2021;48(9):2904–13. https://doi.org/10.1007/s00259-021-05220-7.
26. Berek JS, Matias-Guiu X, Creutzberg C, et al. FIGO staging of endometrial cancer: 2023. J Gynecol Oncol. 2023;34(5):e85. https://doi.org/10.3802/jgo.2023.34.e85.
27. Shifeng T, Yue W, Wen Z, et al. The value of multimodal functional magnetic resonance imaging in differentiating p53abn from p53wt endometrial carcinoma. Acta Radiol. 2023;2841851231198911. https://doi.org/10.1177/02841851231198911.
28. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-Enhancing Lesions at multiparametric breast MRI. Radiology. 2019;290(2):290–7. https://doi.org/10.1148/radiol.2018181352.
29. Dai M, Liu Y, Hu Y, et al. Combining multiparametric MRI features-based transfer learning and clinical parameters: application of machine learning for the differentiation of uterine sarcomas from atypical leiomyomas. Eur Radiol. 2022;32(11):7988–97. https://doi.org/10.1007/s00330-022-08783-7.
30. Yang X, Liu C, Wang Z, et al. Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI. Med Image Anal. 2017;42:212–27. https://doi.org/10.1016/j.media.2017.08.006.
31. Zhang L, Wu X, Liu J, et al. MRI-Based deep-learning model for distant metastasis-free survival in Locoregionally Advanced Nasopharyngeal Carcinoma. J Magn Reson Imaging. 2021;53(1):167–78. https://doi.org/10.1002/jmri.27308.
32. Meng N, Feng P, Yu X, et al. An [18F]FDG PET/3D-ultrashort echo time MRI-based radiomics model established by machine learning facilitates preoperative assessment of lymph node status in non-small cell lung cancer. Eur Radiol. 2023. https://doi.org/10.1007/s00330-023-09978-2.
33. Hong JH, Jung JY, Jo A, et al. Development and validation of a Radiomics Model for differentiating Bone islands and Osteoblastic Bone metastases at Abdominal CT. Radiology. 2021;299(3):626–32. https://doi.org/10.1148/radiol.2021203783.
34. Che D, Liu Q, Rasheed K, Tao X. Decision tree and ensemble learning algorithms with their applications in bioinformatics. Adv Exp Med Biol. 2011;696:191–9. https://doi.org/10.1007/978-1-4419-7046-6_19.
35. Seeger M. Gaussian processes for machine learning. Int J Neural Syst. 2004;14(2):69–106. https://doi.org/10.1142/S0129065704001899.
36. Chen C, Qin Y, Chen H, et al. Machine learning to differentiate small round cell malignant tumors and non-small round cell malignant tumors of the nasal and paranasal sinuses using apparent diffusion coefficient values. Eur Radiol. 2022;32(6):3819–29. https://doi.org/10.1007/s00330-021-08465-w.

Publisher's note