BioMedical Engineering OnLine

RESEARCH

Open Access

Deep learning based on ultrasound images to predict platinum resistance in patients with epithelial ovarian cancer

Chang Su[1], Kuo Miao[1], Liwei Zhang[1] and Xiaoqiu Dong[1*]

*Correspondence:
dongxq0451@163.com

[1] Department of Ultrasound, Fourth Affiliated Hospital of Harbin Medical University, 37 Yi Yuan Street, Harbin 150086, China

Abstract

Background: The study aimed at developing and validating a deep learning (DL) model based on the ultrasound imaging for predicting the platinum resistance of patients with epithelial ovarian cancer (EOC).

Methods: 392 patients were enrolled in this retrospective study who had been diagnosed with EOC between 2014 and 2020 and underwent pelvic ultrasound before initial treatment. A DL model was developed to predict patients' platinum resistance, and the model underwent evaluation through receiver-operating characteristic (ROC) curves, decision curve analysis (DCA), and calibration curve.

Results: The ROC curves showed that the area under the curve (AUC) of the DL model for predicting patients' platinum resistance in the internal and external test sets were 0.86 (95% CI 0.83–0.90) and 0.86 (95% CI 0.84–0.89), respectively. The model demonstrated high clinical value through clinical decision curve analysis and exhibited good calibration efficiency in the training cohort. Kaplan–Meier analyses showed that the model's optimal cutoff value successfully distinguished between patients at high and low risk of recurrence, with hazard ratios of 3.1 (95% CI 2.3–4.1, $P < 0.0001$) and 2.9 (95% CI 2.3–3.9; $P < 0.0001$) in the high-risk group of the internal and external test sets, serving as a prognostic indicator.

Conclusions: The DL model based on ultrasound imaging can predict platinum resistance in patients with EOC and may support clinicians in making the most appropriate treatment decisions.

Keywords: Deep learning, Ultrasound, Epithelial ovarian cancer, Platinum resistance, Platinum free interval

Background

Ovarian cancer (OC) is the third most common and the most lethal gynecologic malignant neoplasm worldwide [1]. Epithelial ovarian cancer (EOC) is the predominant type of OC accounting for 90% of the cases, and treatment of EOC currently relies on surgery, platinum-based chemotherapy, and maintenance therapy with poly ADP-ribose polymerase (PARP) inhibitors [2]. However, approximately 30% of patients

have platinum-resistant ovarian cancer (PROC), with a platinum-free interval (PFI) was less than 6 months [3]. These patients often require other novel treatments such as antibody–drug conjugates, replication stress inhibitors, and immunotherapies [4]. In addition, platinum resistance is a major predictor of patients sensitive to PARP inhibitors [5]. However, such patients often have received several cycles of chemotherapy before platinum resistance is discovered. This delays the optimal treatment timing and severely reduces the quality of patient survival [6].

The mechanisms of platinum resistance are heterogeneous and unclear, and there is currently no effective method to predict platinum resistance in patients [7]. Some studies have used several biomarkers to predict platinum resistance based on biopsies or surgical excisions [8, 9]. However, these methods are invasive and costly, which limits the use of patients. Therefore, approaches predicting platinum resistance in patients with EOC must be developed. Ultrasonography, as a noninvasive method, can clearly show the location, size, blood supply, and other basic characteristics of tumors, making it an ideal method for assessing the prognosis of cancer patients. A study involving 514 patients with OC indicated that machine learning (ML) models based on ultrasound could predict platinum resistance in OC [10]. In addition, ML models based on ultrasound have been developed to predict the prognosis in patients of other cancers [11, 12]. However, ML requires numerous manual operations, which are time consuming. Therefore, new methods are urgently needed to predict platinum resistance in patients with EOC.

Deep learning (DL) is an emerging machine learning technique with powerful algorithms. When compared with traditional ML, DL does not need to extract features manually by the researcher. Hence, processing ultrasound images by DL has a great advantage in predicting platinum resistance in patients with EOC. In the recent years, DL has proven to be a successful approach in predicting the prognosis of patients with cancer. Specifically, the prognosis of liver, breast, rectal, and stomach cancers has been predicted by DL in the recent studies [13–16]. These studies primarily utilize the network structures of convolutional neural networks (CNNs) and vision transformers (VITs). ConvNext network, presented in 2022, was considered an advanced algorithm [17]. However, few studies have employed this algorithm to predict disease prognosis [18, 19].

In this study, a DL model based on the ConvNext algorithm was developed to predict platinum resistance in EOC patients using ultrasound images prior to the primary intervention on the patient.

Results

Characteristics of patients

For the training set, there were 302 patients with EOC, and 86 patients were finally excluded because of the lack of ultrasound images or follow-up data. Similarly, 93 patients with EOC were involved in the test set and 23 patients were excluded due to an absence of ultrasound images or follow-up data. For the external test set, 106 patients were eventually enrolled and 85 patients were excluded. After removing the low-quality images, there were 2060 images of 392 patients with EOC were enrolled in the study (Fig. 1). The characteristics of patients were summarized in Table 1, including

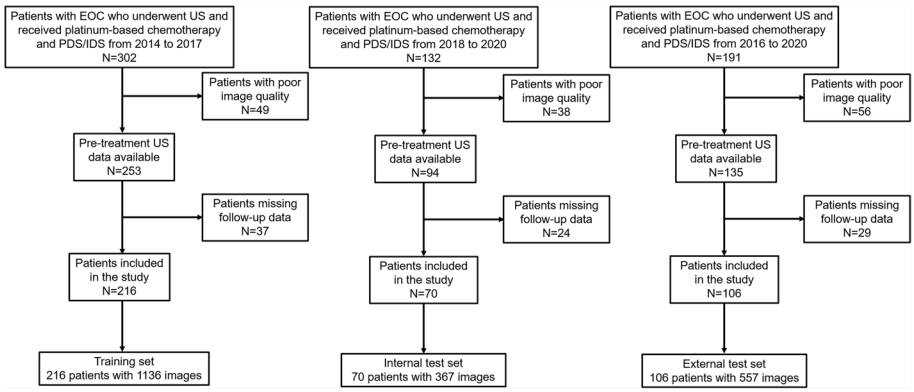

Fig. 1 Flowchart for screening patients. EOC, epithelial ovarian cancer; IDS, interval debulking surgery; PDS, primary debulking surgery; US, ultrasound

the training set (n = 216), internal test set (n = 70), and external test set (n = 106). The median age in the training, internal and external test sets were 58 years (IQR, 51–64 years), 59 years (IQR, 52–62 years), 58 years (IQR, 51–61 years), respectively. There was no statistically significant difference observed in basic characteristics between all the sets.

DL models predict platinum resistance

A ConvNext model was trained to predict patients'platinum resistance. The cross-validation performance for the model was shown in Fig. 2a. As shown in Table 2, the DL model demonstrated a higher area under the receiver-operating characteristic (ROC) curve (AUC) of 0.87 ± 0.01 than the other neural networks in the training set, indicating that the model had good generalization performance. The accuracy and loss plots during training and validation phases were shown in Fig. S1. The result confirmed that the DL model did not overfit or underfit the training data.

The ROC curve analysis results showed that the model could predict patients'platinum resistance in the internal test and external test sets with AUCs of 0.86 (95% CI: 0.83–0.90) and 0.86 (95% CI: 0.84–0.89), respectively (Fig. 2b). The other performance metrics for ConvNext model on the internal and external test sets were shown in Table 3.

Figure 2c, d showed the Calibration curves of the DL model in the two test sets. The curves revealed that the prediction of platinum resistance by the model was accurate. The Hosmer–Lemeshow test was conducted in the two test sets, and the results were X-squared = 11.857, P = 0.158 (P > 0.05) and X-squared = 14.397, P = 0.072 (P > 0.05) in the internal and external test sets, respectively. The results imply that no significant difference was found between the predicted outcome and the actual outcome in both sets. The Calibration curve of the DL model in the training set was shown in Fig. S2a.

Figure 2e and Fig. 2f showed the decision curves that were created to evaluate the effectiveness of models in the internal and external test sets. The results indicated that the DL model could provide higher overall net benefits across most risk thresholds in discriminating platinum resistance and platinum sensitivity of EOC patients. The decision curve of the DL model in the training set was shown in Fig. S2b.

Table 1 Patient Characteristics in the Training, Internal Test, and External Test Sets

Variables	Training set ($n=216$)		*P*	Internal test set ($n=70$)		*P*	External test set ($n=106$)		*P*
	Resistant ($n=64$)	Sensitive ($n=152$)		Resistant ($n=18$)	Sensitive ($n=52$)		Resistant ($n=39$)	Sensitive ($n=67$)	
BMI, mean (SD)	24.12 (2.62)	23.77 (2.82)	0.494	24.11 (3.08)	24.16 (4.21)	0.962	24.07 (1.84)	23.67 (2.72)	0.716
Age, mean (SD), year	56.05 (7.92)	56.44 (8.32)	0.650	56.39 (7.39)	57.88 (7.19)	0.453	57.29 (8.13)	57.54 (8.29)	0.823
CA125, mean (SD), ng/ml	1592.48 (1351.93)	1440.51 (1243.02)	0.486	1327.58 (1437.53)	1223.59 (1113.01)	0.753	1337.68 (1303.16)	1204.37 (1058.11)	0.689
PFI, mean (SD), month	4.52 (1.79)	18.95 (13.02)	<.001	4.39 (1.88)	21.52 (16.87)	<.001	4.74 (1.72)	22.35 (14.36)	<.001
Type of surgery, No. (%)			0.748			0.935			0.951
PDS	21 (32.81)	46 (30.26)		3 (16.67)	7 (13.46)		7 (17.95)	13 (19.40)	
IDS	43 (67.19)	106 (69.74)		15 (83.33)	45 (86.54)		32 (82.05)	54 (80.60)	
HIPEC, No. (%)			0.695			0.211			0.842
No	29 (45.31)	68 (44.74)		14 (77.78)	32 (61.54)		23 (58.97)	38 (56.72)	
Yes	35 (54.69)	84 (55.26)		4 (22.22)	20 (38.46)		16 (41.03)	29 (43.28)	
Histologic classification, *n* (%)			0.349			0.624			0.481
High grade serous carcinoma	59 (92.19)	121 (79.61)		16 (88.89)	45 (86.54)		34 (87.18)	57 (85.07)	
Endometrioid carcinoma	0 (0.00)	6 (3.95)		0 (0.00)	1 (1.92)		0 (0.00)	2 (2.99)	
Clear cell carcinoma	1 (1.56)	3 (1.97)		0 (0.00)	0 (0.00)		0 (0.00)	1 (1.49)	
Mucinous carcinoma	0 (0.00)	2 (1.32)		0 (0.00)	1 (1.92)		1 (2.56)	0 (0.00)	
Low grade serous carcinoma	4 (6.25)	20 (13.15)		2 (11.11)	5 (9.62)		4 (10.26)	7 (10.45)	
FIGO stage, *n* (%)			0.858			0.820			0.871
II	4 (6.25)	6 (3.95)		1 (5.56)	2 (3.85)		1 (2.56)	2 (2.99)	
III	54 (84.38)	127 (83.55)		15 (83.33)	47 (90.38)		33 (84.62)	54 (80.60)	
IV	6 (9.37)	19 (12.50)		2 (11.11)	3 (5.77)		5 (12.82)	11 (16.42)	

BMI, body mass index; CA125, cancer antigen 125; PFI, platinum-free interval; HIPEC, hyperthermic intraperitoneal chemotherapy; FIGO, International Federation of Gynecology and Obstetrics; SD, standard deviation; PDS, primary debulking surgery; IDS, interval debulking surgery

In order to evaluate the robustness of the deep learning model, a subgroup analysis was conducted to assess its performance across different ultrasound systems. The results indicated that the model demonstrated strong predictive performance, regardless of the ultrasound system employed (Supplement Table 1). In addition, the model was retested using noisy datasets (i.e., the ultrasound images with no cropping) to evaluate the model's robustness against image noise. As shown in Supplement Table S2, the model's accuracy on the two test sets was 78.57% and

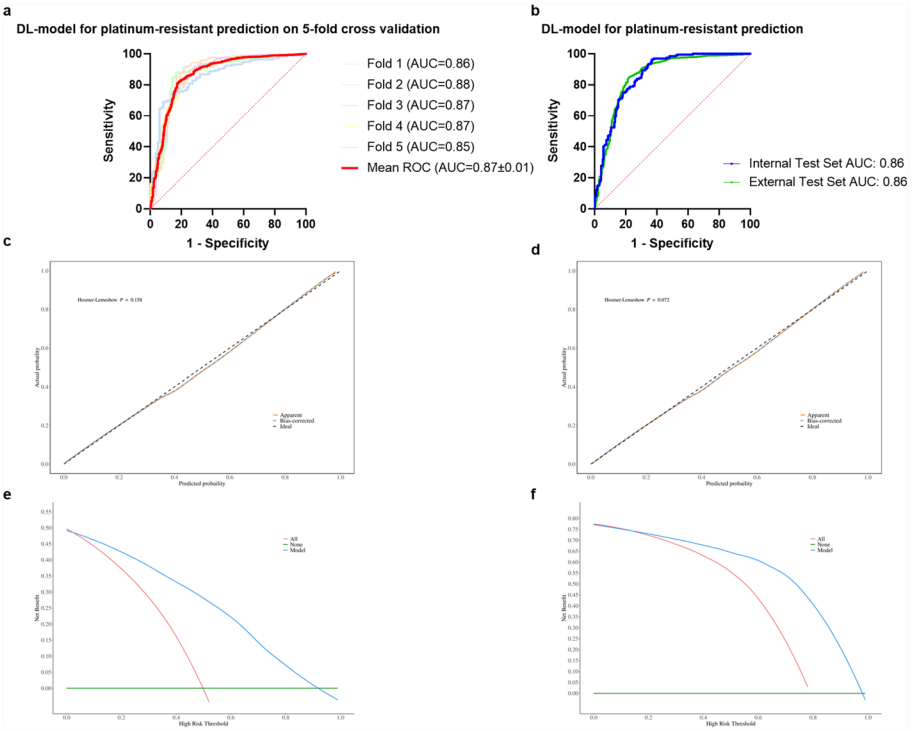

Fig. 2 **a** Receiver-operating characteristic (ROC) curves of deep learning (DL) model performance for predicting platinum resistance on fivefold cross validation. **b** ROC curves of the DL model for platinum-resistant prediction on test sets. **c** Calibration curves for the DL model in the internal test cohort. **d** Calibration curves for the DL model in the external test cohort. (e) Clinical decision curves for the DL model in the internal test cohort. **f** Clinical decision curves for the DL model in the external test cohort

Table 2 Diagnostic Performances of Different Network Structures on 5-Fold Cross-Validation

AUCs	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean AUC
ResNet34	0.86	0.68	0.78	0.88	0.71	0.78 ± 0.08
DenseNet121	0.86	0.88	0.70	0.85	0.78	0.81 ± 0.07
Swin Transformer	0.72	0.67	0.81	0.81	0.73	0.75 ± 0.05
ConvNext	0.86	0.88	0.87	0.87	0.85	0.87 ± 0.01

AUC, area under the receiver operating characteristic curve

Table 3 Diagnostic Performance of the Deep Learning Model in Training, Internal Test and External Test Sets

Sets	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC (95% CI)	F1 score	MCC
Training set	81.94 ± 0.01	82.24 ± 0.01	81.47 ± 0.01	82.81 ± 0.01	83.55 ± 0.01	0.87 ± 0.01	0.75 ± 0.01	0.63 ± 0.01
Internal test set	81.43	82.83	74.00	83.33	80.77	0.86(0.83–0.90)	0.70	0.59
External test set	80.19	85.81	76.36	79.49	80.60	0.86(0.84–0.89)	0.75	0.59

Notes: AUC, area under the receiver-operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; MCC, matthews correlation coefficient

78.30%, respectively, which showed no statistically significant difference compared to the results before adding noise ($P = 0.673$; $P = 0.735$, respectively).

DL model predicts progression-free survival

Kaplan–Meier analysis further demonstrated a high correlation between the decision function value of the DL model and progression-free survival (PFS). Patients were separated into the low-risk and high-risk groups utilizing the optimal cutoff value obtained in the training set (0.7909). As is shown in Fig. 3, the high-risk patients had lower PFS than the low-risk patients in the training set (HR, 2.4 [95% CI: 2.1–2.7]; $P < 0.0001$), internal test set (HR, 3.1 [95% CI: 2.3–4.1]; $P < 0.0001$), and external test set (HR, 2.9 [95% CI: 2.3–3.9]; $P < 0.0001$).

To better validate the DL model performance, the study utilized the model to make predictions for 1-year and 2-year PFS. According to the ROC curves, the AUCs were 0.77, 0.76, and 0.76 in 1-year PFS prediction in the training, internal test, and external test sets, respectively. However, the AUCs were reduced to 0.73, 0.73, and 0.72 for the 2-year prediction for PFS in the same sets (Fig.S3).

A subgroup analysis was conducted on the internal test set of patients. The analysis demonstrated that the model performed well in almost all clinical subgroups (Fig.S4). In the test cohort and subgroups, the use of hyperthermic intraperitoneal chemotherapy (HIPEC) could improve PFS, even in the high-risk group which is generally considered to have a worse prognosis (HR, 2.0 [95% CI: 1.3–3.0]; $P = 0.0004$) (Fig.S5).

As shown in Fig.S6, the HRs of the DL model-predicted risk, adjusted by multivariable Cox regression, were 8.2 (95% CI: 5.4–12.3; $P < 0.0001$), 4.6 (95% CI: 3.1–6.8; $P < 0.0001$), and 3.9 (95% CI: 2.8–6.3; $P < 0.0001$) in the training, internal test, and external test sets, respectively. The HRs of the DL model-predicted risk were higher than the other factors, which showed that the DL model prediction was statistically independent of potential confounders.

The example Grad-CAMs and their pre-treatment ultrasound images are shown in Fig. 4. In Grad-CAMs, colors closer to red and blue represent higher- and lower-weighted regions in the network, respectively. As shown in the figure, patients A and B were high-risk group patients with a shorter PFS and a lower decision function value due to platinum resistance. In contrast, patients C and D were low-risk group patients with a longer PFS and a higher decision function value due to platinum sensitivity. In grad-cam heatmaps, we found that the high-response areas seem to be the location of blood vessels according to color Doppler ultrasound images shown in Fig. 4. We used

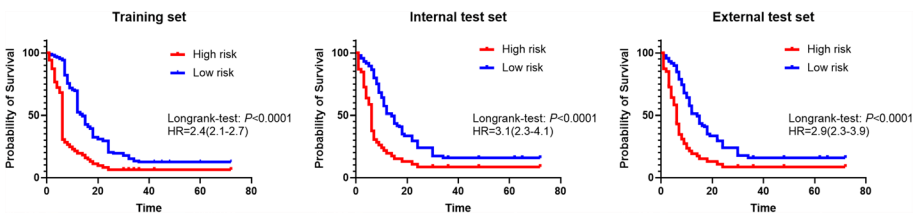

Fig. 3 Kaplan–Meier curves of progression-free survival (PFS) according to the decision function value of deep learning (DL) model in patients split into high-risk and low-risk groups from the training, internal test and external test sets

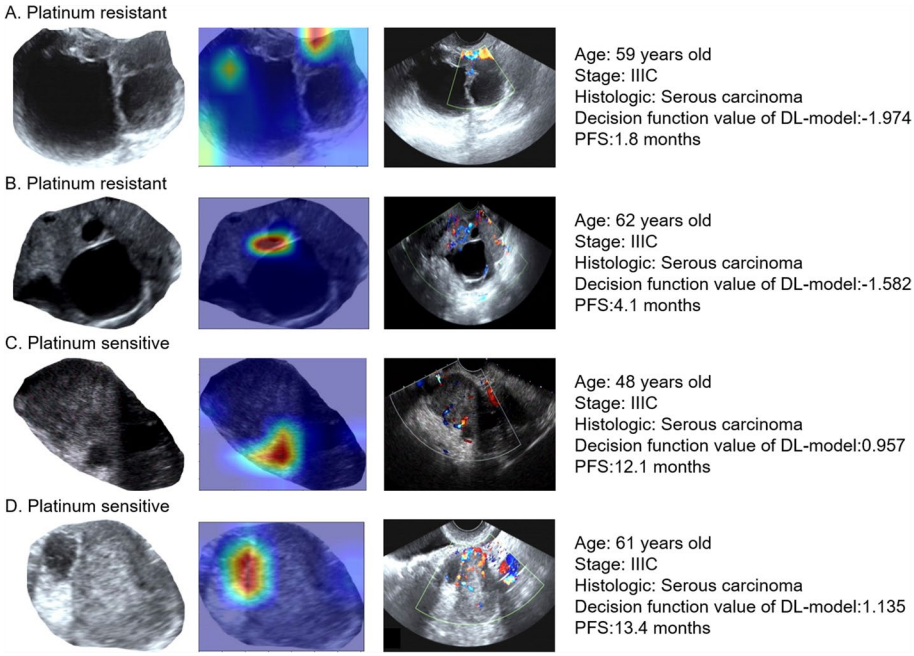

Fig. 4 Representative prediction results of the deep learning (DL) model. The gradient-weighted class activation maps (Grad-CAM) and color Doppler ultrasound images are shown on the right side of the input image

the Dice score to calculate the consistency between the red regions in the Grad-CAM map and the blood vessel regions annotated by the physician. The blood vessel regions were annotated by a physician with 10 years of gynecological ultrasound experience and reviewed by a physician with 38 years of gynecological ultrasound experience. The results showed that the Dice score was 0.7990 and 0.7866 in the internal and external test sets, respectively, indicating good consistency between high-response and blood vessel areas (Fig. S7).

Discussion

In this study, we developed a DL model based on ConvNext for predicting platinum resistance in patients with EOC from the ultrasound images, which showed considerable predictive performance in the training set and two test sets. The model predicted platinum resistance in EOC patients and had a high specificity and sensitivity. The calibration curve and decision curve analyses showed that the DL model displayed optimal calibration and provided significant clinical benefits, highlighting its potential as a trustworthy tool in clinical decision making. To assess the model overall, the Matthews correlation coefficient (MCC) was calculated in both test sets. The results confirmed that the model performed well in predicting both the positive and negative classes [20]. Moreover, good results were obtained in the prediction of one-year survival of patients by this model. Cox regression analysis showed that the decision function value based on the DL model was an independent factor affecting EOC. The RadImageNet dataset for model pretraining and the standardized data preprocessing procedure provided in the study have also enhanced the overall performance and applicability of the model.

Platinum resistance is the leading cause of death and reduced PFS in OC patients, however, the mechanism of platinum resistance is not fully understood. In this study, the ultrasound images were applied to predict the platinum resistance of patients with EOC, which is relatively inexpensive and radiation-free as compared to the studies based on MR or CT [21, 22]. Ultrasonography could also provide real time, dynamic, and multiangle visualization of the lesions, which has great advantages for analyzing the prognosis of EOC patients. Artificial intelligence (AI) has made great progress in the prediction of disease prognosis in the recent years. Chen et al. utilized ML to predict platinum resistance and PFS in EOC with good results [23]. Yang et al. developed an ML model based on clinical data to predict platinum-resistant recurrence of EOC [24]. However, traditional machine learning methods required manual extraction and analysis of features, which greatly increased the workload. Deep learning, as a special form of machine learning, could automatically extract features from images with improved accuracy. A DL model based on ultrasound for predicting the resistance to neoadjuvant chemotherapy in breast cancer had been built and achieved satisfactory results [25]. However, few studies have applied DL for drug resistance prediction in EOC. The ROC curves, decision and calibration curves verified the good performance of this DL model.

In the study, we used ConvNext as the network backbone, which is a modernized version of the standard ResNet50 that combines the advantages of CNN and ViT for a simpler structure and better training [26]. There have been studies in the past using CNNs or ViTs to predict the prognosis of EOC [27, 28]. However, there are relatively few studies applying ConvNext to medical image recognition as well as tumor prognosis prediction. In the study, the prediction efficacy of ConvNext was compared with ResNet34, DenseNet121, and Swin Transformer on the training set using ROC curves, and the results indicated that the AUC of ConvNext exceeded the other three in the training set, which proved its good performance. Compared with the traditional CNN networks, ConvNext adjusts the number and ratio of blocks on the macro level and adjusts the ReLU activation function to the GELU activation function on the micro level, all of these adjustments lead to the improvement of the model's training effectiveness. As a type of ViT, Swin Transformer also performed well in image recognition, however, previous studies have demonstrated that Swin Transformer's performance is better when the image resolution is 384×384 [29]. The increase in resolution leads to a greater amount of computation, and it even requires a TPU environment to complete the computation. Also, Swin Transformer's network structure is more complex than ConvNext, which will greatly improve the training time as well as the training cost. Therefore, the ConvNext model has greater potential in the area of image recognition.

Pretrained weights from the RadImageNet models were utilized as the initial basis for the DL model in this study. The RadImageNet database included ultrasound images, such as pelvic ultrasound images, that closely resembled our EOC data [30]. This similarity contributed to an enhancement in the performance of the DL model. The consistency of DL model performance across various ultrasound devices is crucial for their practical application in real-world situations. In this study, we implemented a standardized image preprocessing procedure to eliminate discrepancies between devices and validated the model's stability through subgroup analysis. This approach aligned with the method used by Christiansen et al., who employed DL to differentiate between benign and malignant

ovarian tumors, suggesting that standardized image preprocessing can improve the robustness of DL models [31]. The imbalance in the datasets was addressed by data augmentation in the study. Hu et al. used an imaging oversampling method to achieve a balance of the dataset, which may not alter the feature distribution of the original data [32]. We will explore this method to better enhance the generalization performance of the model in future studies.

In this study, the capacity of the model to predict PFS in patients with EOC was validated using Kaplan–Meier analysis and further validated in subgroup analysis. Cox regression analysis also demonstrated that the model's prediction of patients'PFS was statistically independent of other confounders. As indicated by the ROC curves, the DL model showed good predictive power for 1-year PFS but performed poorly in predicting 2-year PFS. Jiang et al. used a ViT model to predict the overall survival (OS) of rectal cancer patients, and the results showed that the model's AUC for predicting one-year, three-year, and five-year OS were 0.87, 0.54, and 0.65, respectively [15]. Lei et al. used a ResNet model to predict the PFS of ovarian cancer patients. The model achieved an AUC of 0.98 for one-year PFS prediction and dropped to 0.71 for two-year PFS prediction [21]. Similar to these results, the model in this study also experienced a significant decline in performance when predicting long-term survival. Therefore, more complex models may be needed for predicting long-term survival. Mei et al. constructed a long-short-term memory (LSTM) model to predict the three-year survival rate of patients with interstitial lung disease [33]. If this model could be applied here, it might be able to address the current issues in this study and achieve the prediction of long-term survival rates. In addition, the emergence of more confounding factors (patient's economic status, treatment adherence, etc.) with the passing of follow-up time also influenced the accuracy of the model. In future studies, we will include more patients and increase the number of centers to enhance the diversity of the data. Overall, the decision function value generated by ConvNext-based deep learning model was a good short-term prognostic indicator but weakened in predicting long-term outcomes. The HIPEC is a significant treatment option for EOC, and previous studies have demonstrated that the HIPEC significantly reduces recurrence and death and prolongs median survival time in OC [34]. This study also validated this using Kaplan–Meier analysis, and the use of HIPEC significantly improved PFS in both high- and low-risk patient cohorts predicted by the deep learning model. This also informs therapeutic decision-making for patients with EOC.

Deep learning has been made a 'black box' due to its features that are often difficult to interpret, so visualizing these features has huge benefits for people to understand the decision logic of deep learning [35]. In this study, we used grad-cam heatmaps to make observations about deep learning analytics. In grad-cam heatmaps, we identified the correlation between high-response regions and blood vessels and validated it using quantitative analysis methods. In the previous studies, the high number of neovessels and their heterogeneous distribution may be an important reason for the increase in tumor heterogeneity [36]. Therefore, the formation of tumor neovascularization might be closely related to the patient's resistance to chemotherapeutic agents.

There are some limitations of this study. Firstly, this study is a double-center study, and there is a lack of a larger sample to further validate the robustness of the model. Secondly,

this study is retrospective and lacks some data at the genetic level, more prospective studies will be conducted in the future and will further explain the characteristics of deep learning at the genetic level. Thirdly, although standard image preprocessing was performed, we will continue to explore domain adaptation techniques in future studies to better eliminate the impact of differences between ultrasound devices on DL models. Lastly, although the DL model showed good results in predicting platinum resistance, it is unclear which images to store may affect the model's performance as this study was conducted retrospectively. We will conduct further research to clarify the impact of image selection on model performance.

Conclusion

This study demonstrates that a deep learning model based on ultrasound images can serve as an effective biomarker for predicting platinum resistance and 1-year PFS in EOC patients. The model eliminates the need for manual feature extraction, providing a more personalized and accurate patient assessment. In addition, the study highlights the potential superiority of the ConvNext-based neural network model over CNNs and ViTs in medical image analysis, offering valuable insights for clinical decision making.

Methods

Patients

This retrospective study was performed following the Declaration of Helsinki. Three independent patient cohorts were collected in this study. Initially, patients diagnosed with EOC at the Fourth Affiliated Hospital of Harbin Medical University from 2014 to 2017 were randomly assigned to either the training or validation set. Subsequently, patients diagnosed with EOC at the same center from 2018 to 2020 were split into the internal test set. Lastly, the external test set included patients with EOC at Harbin New Area Central Hospital from 2016 to 2020 (Fig. 1). The inclusion criteria were: (1) pathologically confirmed EOC based on surgical resection; (2) pelvic ultrasound examination available within 1 week before initial treatment; (3) underwent cisplatin/carboplatin chemotherapy for at least four cycles during the initial treatment or changed to a different chemotherapy regimen because of primary platinum-refractory status; (4) did not receive additional treatment in the first two years after diagnosis. The exclusion criteria were: (1) pelvic ultrasound examination not performed before initial treatment or the quality of the images was not sufficient; (2) either missing or incomplete data available for follow-up; (3) combination with other malignant tumors.

Data collection

Gray-scale pelvic ultrasound images of patients before initial treatment were acquired using transvaginal ultrasonography and combined with transabdominal ultrasonography if the patient had a large mass that could not be adequately assessed by transvaginal ultrasonography. Ultrasound images of the lesion were acquired from multiple angles, including the plane of the maximum diameter tumor and its vertical plane, the plane of maximum diameter of the tumor solid portion and its vertical plane, and the planes featuring key lesion characteristics such as irregular internal wall or papillary projections. Five to six ultrasound images were collected from each patient. The

ultrasound diagnostic systems used in the study included Samsung WS80 A, Philips EPIQ7, GE Voluson S8, and Siemens Acuson S3000. For the quality control of ultrasound images, we checked all images to remove those with low quality (Fig. 1). The tumor's areas of interest were delineated manually by a sonographer with more than 3 years of gynecological ultrasound experience, and reviewed by an expert experienced for more than 10 years in EOC. The detailed process was described in the Appendix Methods.

The hospital electronic analysis system was used to obtain the follow-up data, including age, body mass index (BMI), differentiation stage, histological type, International Federation of Gynecology and Obstetrics (FIGO) stage, preoperative cancer antigen 125 (CA125), surgical procedures, and HIPEC or not.

Patients had a follow-up in accordance with NCCN guidelines. The last follow-up date was December 31, 2023. The first follow-up endpoint was the PFI. Specifically, patients whose PFI was less than six months were classified to the platinum-resistant group. Instead, the patients with a PFI of six months or more than six months were classified to the platinum-sensitive group. The secondary follow-up endpoint was PFS. PFS was calculated from the date of primary treatment to the date of recurrence or to the last no-recurrence date for those who were censored. Recurrence includes biochemical recurrence, clinical recurrence, and imaging recurrence.

Image preprocessing

Before training, the regions of interest within the tumors were cropped on ultrasound images of all the three cohorts (Fig. 5a). To enhance the training process, the data augmentations were employed in the training set including random cropping (224 $\times$224 pixels), random horizontal flipping ($p=0.5$), random rotation ($-15°$ to $+15°$), and color jittering (brightness $=0.5$, contrast $=0.5$, saturation $=0.5$). This was done to improve the model's ability to generalize effectively. The images in the two test sets were

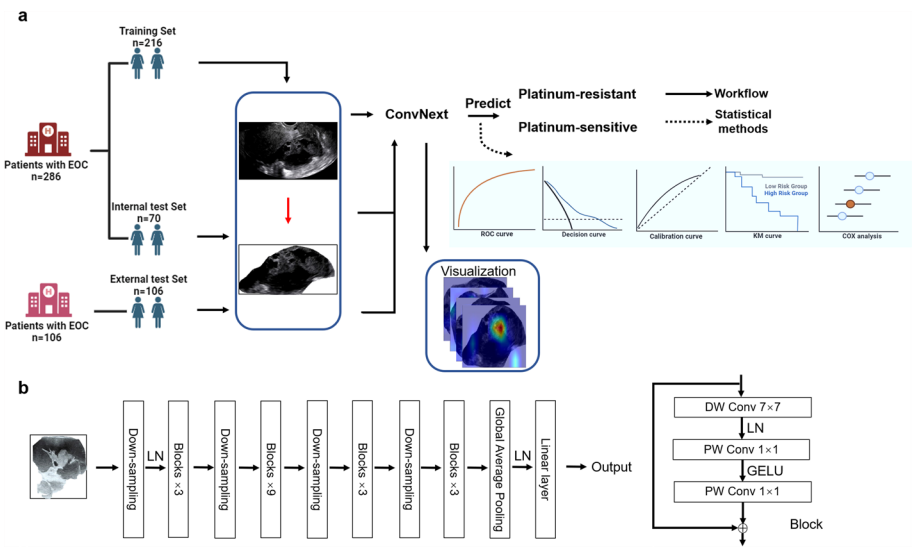

Fig. 5 Overview of the workflow of this study. **a** Diagram shows the workflow for the development and validation of the deep learning (DL) model. Icons reproduced from https://biorender.com. **b** Diagram shows the model structure. DW Conv: Depthwise Convolution, PW Conv: Pointwise Convolution

not augmented to prevent bias in the results. The images in all three cohorts were then normalized to a range of 0 to 1 intensity and adjusted to 224 × 224 pixels, conforming to the standard input dimensions of our baseline DL architectures. The image preprocessing process was applied to enhance data quality and consistency and balance the number of images of the platinum-resistant group and the platinum-sensitive group to improve the model's generalizability and accuracy. The number of images in the training set after image preprocessing was shown in Supplement Table 3.

Model building and validation

We developed a ConvNext-based deep learning model to predict a patient's platinum resistance from grayscale ultrasound images [17]. The structure of the model was shown in Fig. 5b. The initial weights of the ConvNext model were derived from a pretrained model from RadImageNet [30]. 50 epochs in training session were performed using the Adam optimizer (initialized with the learning rate set to 0.0001). A dropout layer at a rate of 0.5 was added to mitigate overfitting. The objective function was a weighted cross-entropy loss, which assigned a decision function value to each image. In building the model, the fivefold cross-validation was used to evaluate the model performance. The dataset splitting was done based on the number of patients. Consequently, images from the same patient were all in the same dataset. The predicted value for the lesion was determined as the average of its decision function values for all images of that lesion [37]. A gradient-weighted class activation mapping (Grad-CAM) was performed to visualize the essential parts of images for prediction. In addition, we compared the model's performance with other deep learning models including ResNet34 [38], DenseNet121 [39], and Swin Transformer [40]. All the programs were executed using Python version 3.6.8.

Statistical analysis

R, version 4.3.0 (R Foundation) and GraphPad Prism 9 (San Diego, CA, USA) were utilized for data analysis. The Chi-square or Fisher exact test was performed for categorical variable comparisons, and the t test was used for analyzing continuous variables. A ROC curve analysis was utilized to establish the optimal cutoff value that maximizes prediction accuracy. The AUCs were compared using the DeLong test. Model performance was assessed by area under the curve, positive predictive value, negative predictive value, sensitivity, specificity, F1 score, and MCC. Moreover, the model was also evaluated by a Hosmer–Lemeshow test, a calibration curve and a decision curve analysis (DCA). Based on the optimal threshold determined in the training set, each patient cohort was split into low-risk and high-risk groups. Kaplan–Meier survival analysis was utilized to evaluate the prognosis difference between patients in the two groups. Subgroup analysis was performed to investigate the relationship between the two groups, which were divided by age, BMI, pretreatment CA125 level (in ng/ml), and HIPEC. A Cox regression analysis was conducted to calculate the log-rank test, hazard ratios (HRs), and 95% confidence intervals (CIs). A P value < 0.05 was used to define significant results.

Abbreviations

AUC	Area under the ROC curve
BMI	Body Mass Index
CA125	Cancer antigen 125
CI	Confidence interval
CNN	Convolutional neural networks
DCA	Decision curve analysis
DL	Deep learning
EOC	Epithelial ovarian cancer
FIGO	International Federation of Gynecology and Obstetrics
Grad-CAM	Gradient-weighted class activation mapping
HIPEC	Hyperthermic intraperitoneal chemotherapy
HR	Hazard ratio
LSTM	Long-short-term memory
MCC	Matthews correlation coefficient
ML	Machine learning
OC	Ovarian cancer
OS	Overall survival
PARP	Poly ADP-ribose polymerase
PFI	Platinum-free interval
PFS	Progression-free survival
PROC	Platinum-resistant ovarian cancer
ROC	Receiver-operating characteristic
VIT	Vision transformers

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12938-025-01391-8.

Additional file 1.

Acknowledgements
The authors thank all participants who made valuable contributions to this study, including all the patients and experts of the Department of Ultrasound, Fourth Affiliated Hospital of Harbin Medical University.

Author contributions
DXQ contributed to the conception and design of the work. SC and MK were involved in data acquisition, analysis, or interpretation. SC and ZLW spearheaded the data processing and deep learning analysis. SC wrote the main manuscript. All authors reviewed the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding
This work was supported by the Natural Science Foundation of Heilongjiang Province, China (Grant No. LH2022H033) and the Research Program Da'ai Longjing Charity Foundation of Heilongjiang Province (Grant No. HX2020-20).

Availability of data and materials
The data that support the findings of this study were available upon request from the corresponding author. The data were not publicly available due to privacy or ethical restrictions.

Declarations

Ethics approval
This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee the Fourth Affiliated Hospital of Harbin Medical University.

Consent to participate
Written informed consent was waived by the Institutional Review Board of the Fourth Affiliated Hospital of Harbin Medical University.

Consent to publication
The authors affirm that human research participants provided informed consent for publication of the images in Figs. 4 and 5.

Competing interests
The authors declare no competing interests.

Received: 6 January 2025 Accepted: 2 May 2025
Published online: 13 May 2025

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72:7–33.
2. Liu J, Berchuck A, Backes FJ, Cohen J, Grisham R, Leath CA, et al. NCCN guidelines(R) insights: ovarian cancer/fallopian tube cancer/primary peritoneal cancer, version 3.2024. J Natl Compr Canc Netw. 2024;22:512–9.
3. Vizzielli G, Giudice MT, Nardelli F, Costantini B, Salutari V, Inzani FS, et al. Pressurized IntraPeritoneal Aerosol Chemotherapy (PIPAC) Applied to Platinum-Resistant recurrence of ovarian tumor: a single-institution experience (ID: PARROT Trial). Ann Surg Oncol. 2024;31:1207–16.
4. Richardson DL, Eskander RN, O'Malley DM. Advances in ovarian cancer care and unmet treatment needs for patients with platinum resistance: a narrative review. JAMA Oncol. 2023;9:851–9.
5. Ahn B, Moon D, Kim HS, Lee C, Cho NH, Choi HK, et al. Histopathologic image-based deep learning classifier for predicting platinum-based treatment responses in high-grade serous ovarian cancer. Nat Commun. 2024;15:4253.
6. Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E, et al. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. Nat Commun. 2019;10:764.
7. Yang L, Xie HJ, Li YY, Wang X, Liu XX, Mai J. Molecular mechanisms of platinum-based chemotherapy resistance in ovarian cancer (Review). Oncol Rep. 2022. https://doi.org/10.3892/or.2022.8293.
8. Atallah GA, Kampan NC, Chew KT, Mohd Mokhtar N, Md Zin RR, Shafiee MNB, et al. Predicting Prognosis and Platinum Resistance in Ovarian Cancer: Role of Immunohistochemistry Biomarkers. Int J Mol Sci. 2023. https://doi.org/10.3390/ijms24031973.
9. Wu X, Shen S, Qin J, Fei W, Fan F, Gu J, et al. High co-expression of SLC7A11 and GPX4 as a predictor of platinum resistance and poor prognosis in patients with epithelial ovarian cancer. BJOG. 2022;129(Suppl 2):40–9.
10. Zuo R, Li X, Hu J, Wang W, Lu B, Zhang H, et al. Prediction of ovarian cancer prognosis using statistical radiomic features of ultrasound images. Phys Med Biol. 2024. https://doi.org/10.1088/1361-6560/ad4a02.
11. Cui H, Sun Y, Zhao D, Zhang X, Kong H, Hu N, et al. Radiogenomic analysis of prediction HER2 status in breast cancer by linking ultrasound radiomic feature module with biological functions. J Transl Med. 2023;21:44.
12. Barteselli C, Mazza S, Ravetta V, Viera FT, Veronese L, Frigerio C, et al. Ultrasound patterns of hepatocellular carcinoma and their prognostic impact: a retrospective study. Cancers (Basel). 2023. https://doi.org/10.3390/cancers15225396.
13. Jia W, Shi W, Yao Q, Mao Z, Chen C, Fan A, et al. Identifying immune infiltration by deep learning to assess the prognosis of patients with hepatocellular carcinoma. J Cancer Res Clin Oncol. 2023;149:12621–35.
14. Zhang J, Wu J, Zhou XS, Shi F, Shen D. Recent advancements in artificial intelligence for breast cancer: Image augmentation, segmentation, diagnosis, and prognosis approaches. Semin Cancer Biol. 2023;96:11–25.
15. Jiang X, Zhao H, Saldanha OL, Nebelung S, Kuhl C, Amygdalos I, et al. An MRI deep learning model predicts outcome in rectal cancer. Radiology. 2023;307: e222223.
16. Jin C, Jiang Y, Yu H, Wang W, Li B, Chen C, et al. Deep learning analysis of the primary tumour and the prediction of lymph node metastases in gastric cancer. Br J Surg. 2021;108:542–9.
17. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie SJae-p. A ConvNet for the 2020s. 2022.
18. Lu N, Guan X, Zhu J, Li Y, Zhang J. A contrast-enhanced CT-based deep learning system for preoperative prediction of colorectal cancer staging and RAS mutation. Cancers (Basel). 2023. https://doi.org/10.3390/cancers15184497.
19. Zhang L, Wang R, Gao J, Tang Y, Xu X, Kan Y, et al. A novel MRI-based deep learning networks combined with attention mechanism for predicting CDKN2A/B homozygous deletion status in IDH-mutant astrocytoma. Eur Radiol. 2024;34:391–9.
20. Abbasian Ardakani A, Airom O, Khorshidi H, Bureau NJ, Salvi M, Molinari F, et al. Interpretation of Artificial Intelligence Models in Healthcare: A Pictorial Guide for Clinicians. J Ultrasound Med. 2024;43:1789–818.
21. Lei R, Yu Y, Li Q, Yao Q, Wang J, Gao M, et al. Deep learning magnetic resonance imaging predicts platinum sensitivity in patients with epithelial ovarian cancer. Front Oncol. 2022;12: 895177.
22. Wang S, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, et al. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. Radiother Oncol. 2019;132:171–7.
23. Yao F, Ding J, Hu Z, Cai M, Liu J, Huang X, et al. Ultrasound-based radiomics score: a potential biomarker for the prediction of progression-free survival in ovarian epithelial cancer. Abdom Radiol (NY). 2021;46:4936–45.
24. Yang LR, Yang M, Chen LL, Shen YL, He Y, Meng ZT, et al. Machine learning for epithelial ovarian cancer platinum resistance recurrence identification using routine clinical data. Front Oncol. 2024;14:1457294.
25. Gu J, Zhong X, Fang C, Lou W, Fu P, Woodruff HC, et al. Deep learning of multimodal ultrasound: stratifying the response to neoadjuvant chemotherapy in breast cancer before treatment. Oncologist. 2024;29:e187–97.
26. Hassanien MA, Singh VK, Puig D, Abdel-Nasser M. Predicting breast tumor malignancy using deep ConvNeXt radiomics and quality-based score pooling in ultrasound sequences. Diagnostics (Basel). 2022. https://doi.org/10.3390/diagnostics12051053.
27. Wessels F, Schmitt M, Krieghoff-Henning E, Nientiedt M, Waldbillig F, Neuberger M, et al. A self-supervised vision transformer to predict survival from histopathology in renal cell carcinoma. World J Urol. 2023;41:2233–41.
28. Valsaraj A, Kalmady SV, Sharma V, Frost M, Sun W, Sepehrvand N, et al. Development and validation of echocardiography-based machine-learning models to predict mortality. EBioMedicine. 2023;90: 104479.
29. Tummala S, Kadry S, Bukhari SAC, Rauf HT. Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling. Curr Oncol. 2022;29:7498–511.
30. Mei X, Liu Z, Robson PM, Marinelli B, Huang M, Doshi A, et al. RadImageNet: an open radiologic deep learning research dataset for effective transfer learning. Radiol Artif Intell. 2022;4: e210315.
31. Christiansen F, Konuk E, Ganeshan AR, Welch R, Pales Huix J, Czekierdowski A, et al. International multicenter validation of AI-driven ultrasound detection of ovarian cancer. Nat Med. 2025;31:189–96.
32. Hu C, Chen W, Li F, Zhang Y, Yu P, Yang L, et al. Deep learning radio-clinical signatures for predicting neoadjuvant chemotherapy response and prognosis from pretreatment CT images of locally advanced gastric cancer patients. Int J Surg. 2023;109:1980–92.

33. Mei X, Liu Z, Singh A, Lange M, Boddu P, Gong JQX, et al. Interstitial lung disease diagnosis and prognosis using an AI system integrating longitudinal data. Nat Commun. 2023;14:2272.
34. Lei Z, Wang Y, Wang J, Wang K, Tian J, Zhao Y, et al. Evaluation of cytoreductive surgery with or without hyperthermic intraperitoneal chemotherapy for stage III Epithelial ovarian cancer. JAMA Netw Open. 2020;3: e2013940.
35. Ling Y, Ying S, Xu L, Peng Z, Mao X, Chen Z, et al. Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase CT scans with minimum extra information. Front Oncol. 2022;12: 960178.
36. Fu S, Corr BR, Culm-Merdek K, Mockbee C, Youssoufian H, Stagg R, et al. Phase Ib study of Navicixizumab Plus Paclitaxel in patients with platinum-resistant ovarian, primary peritoneal, or fallopian tube cancer. J Clin Oncol. 2022;40:2568–77.
37. Qi Q, Huang X, Zhang Y, Cai S, Liu Z, Qiu T, et al. Ultrasound image-based deep learning to assist in diagnosing gross extrathyroidal extension thyroid cancer: a retrospective multicenter study. EClinicalMedicine. 2023;58: 101905.
38. He K, Zhang X, Ren S, Sun JJI. 2016. Deep Residual Learning for Image Recognition.
39. Huang G, Liu Z, Laurens VDM, Weinberger KQJICS. 2016. Densely Connected Convolutional Networks.
40. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.

Publisher's Note