



OPEN ACCESS

EDITED BY

Sharon R. Pine,
University of Colorado Anschutz Medical
Campus, United States

REVIEWED BY

Martina Arcieri,
Ospedale Santa Maria della Misericordia di
Udine, Italy
Rodrigo Guarischi-Sousa,
Palo Alto Veterans Institute for Research,
United States
Alice Poli,
Ospedale Santa Maria della Misericordia di
Udine, Italy

*CORRESPONDENCE

Nayiyuan Wu

✉ wunayiyuan@163.com

Qianxi Ni

✉ niqianxi@hnca.org.cn

†These authors have contributed equally to
this work

RECEIVED 05 February 2025

ACCEPTED 19 May 2025

PUBLISHED 04 June 2025

CITATION

Xiong X, Cai L, Yang Z, Cao Z, Wu N and Ni Q
(2025) Multimodal data integration
with machine learning for predicting PARP
inhibitor efficacy and prognosis in
ovarian cancer.
Front. Oncol. 15:1571193.
doi: 10.3389/fonc.2025.1571193

COPYRIGHT

© 2025 Xiong, Cai, Yang, Cao, Wu and Ni. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Multimodal data integration with machine learning for predicting PARP inhibitor efficacy and prognosis in ovarian cancer

Xi'an Xiong^{1†}, Li Cai^{1,2†}, Zhen Yang³, Zhongping Cao⁴,
Nayiyuan Wu^{1,5*} and Qianxi Ni^{1*}

¹The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University/Hunan Cancer Hospital, Changsha, China, ²College of Pharmacy, Dali University, Dali, China, ³Oncology Department of Xiangya Hospital, Central South University, Changsha, China, ⁴Department of Oncology of the Second XiangYa Hospital, Central of South University, Changsha, China, ⁵Hunan Key Laboratory Of The Research and Development of Novel Pharmaceutical Preparations, Changsha Medical University, Changsha, China

Background: Poly(ADP)-ribose polymerase inhibitors (PARPi) have brought a significant breakthrough in the maintenance treatment of ovarian cancer. However, beyond BRCA mutation/HRD, the direct impact of other prognostic factors on PARPi response and prognosis remains inadequately characterized.

Methods: We assessed PARPi prognostic factors from clinical characteristics, pathological findings, and biochemical indicators from 251 ovarian cancer patients. Cox univariate and multivariate analyses were employed to identify the factors which influencing PARPi efficacy and patients prognosis. Feature screening was conducted using correlation analysis, significance analysis, Variance Inflation Factor (VIF), and Elastic Net stability analysis. Patient-specific efficacy and prognosis prediction models were then constructed using various machine learning algorithms.

Results: Total bile acids (TBAs) and CA-199 present as an independent risk factor in Cox multivariate analysis for primary and recurrent ovarian cancer patients respectively ($P < 0.05$). TBAs emerged as a risk factor, with each unit increase associated with a 10% rise in recurrence risk. The best-performing model has an AUC of 0.79 ± 0.09 and an AUC of 0.72 ± 0.03 for primary and recurrent ovarian cancer patients respectively. External validation ($n=36$) in multicenter cohorts maintained robust performance with AUC of 0.74 and an AUC of 0.70 for primary and recurrent ovarian cancer patients respectively.

Conclusions: We identified TBAs and CA-199 as a significant prognostic factor in primary and recurrent ovarian cancer patients respectively. The integration of multimodal data with machine learning holds significant potential for enhancing prognosis prediction in PARPi treatment for ovarian cancer.

KEYWORDS

PARP inhibitors (PARPi), prognostic factor, ovarian cancer, machine learning, prediction model

1 Background

Ovarian cancer, often referred to as the “silent killer”, presents the highest mortality rate among gynecological malignancies (1). The standard treatment for ovarian cancer includes cytoreductive surgery followed by systemic platinum-taxane combination chemotherapy (2). Although most patients achieve clinical remission with initial therapy, about 70% of patients may relapse within 2 to 3 years and eventually develop platinum resistance. The 5-year survival rate remains approximately 40% (3–5). The introduction of PARPi has significantly advanced the treatment of ovarian cancer. PARPi can induce apoptosis and death in cancer cells with BRCA mutations or other homologous recombination deficiencies (HRD) through a mechanism known as the “synthetic lethal” effect (6–8). Numerous studies have demonstrated that PARPi can significantly improve progression-free survival (PFS) in patients with BRCA-mutated ovarian cancer (9–12). Consequently, BRCA gene mutations or HRD are commonly utilized as biomarkers for the application of PARPi therapy (13). However, clinical evidence indicates that patients without BRCA mutations or HRD can also derive benefits from PARPi therapy (14–16). And more than 40% patients with BRCA mutations or HRD failed to benefit from PARPi (17, 18). One potential reason for this discrepancy is the insufficient consideration of various prognostic factors in clinical studies evaluating PARPi efficacy in ovarian cancer. These studies often fail to thoroughly investigate which clinicopathological factors might serve as reliable prognostic indicators for PARPi response. Beyond BRCA mutations and HRD, few prognostic factors are currently utilized to guide the clinical application of PARPi. A meta-analysis by Huang et al., encompassing 20 prospective studies, identified BRCA mutation, HRD-positive status, and platinum sensitivity as significant prognostic factors for PARPi efficacy in ovarian cancer; however, other clinicopathological variables did not show a significant predictive value for PFS (19). In contrast, Yusuke et al. demonstrated that HRD status, age, pathological stage, and residual disease status post-cytoreductive surgery are critical prognostic factors for ovarian cancer (20). Bile acids (BAs) have been recognized for their potential role in preventing ovarian cancer by inhibiting proliferation, invasion, and epithelial-mesenchymal transition, as well as enhancing chemotherapy efficacy (21–23). Furthermore, BAs can modulate the expression and activity of multiple PARP enzymes, which may ultimately improve patient survival (24, 25). Lamkin et al. found that higher glucose levels were associated with shorter survival time in ovarian cancer patients in univariate analysis ($HR = 1.88$; $P < 0.05$). Multivariate analysis, adjusted for tumor stage, showed that higher glucose levels were associated with shorter survival time ($HR = 2.01$; $P = 0.04$) and disease-free interval ($HR = 2.32$; $P < 0.05$) (26). Additionally, Zhu et al. identified elevated postoperative CA-199 as an independent risk factor for both PFS and overall survival (OS) in patients with normal postoperative CA-125 levels. The combination of postoperative CA-199 and CA-125 levels offers significant prognostic value for patients with ovarian clear cell carcinoma following initial debulking surgery (27). Recently, Taliento et al.

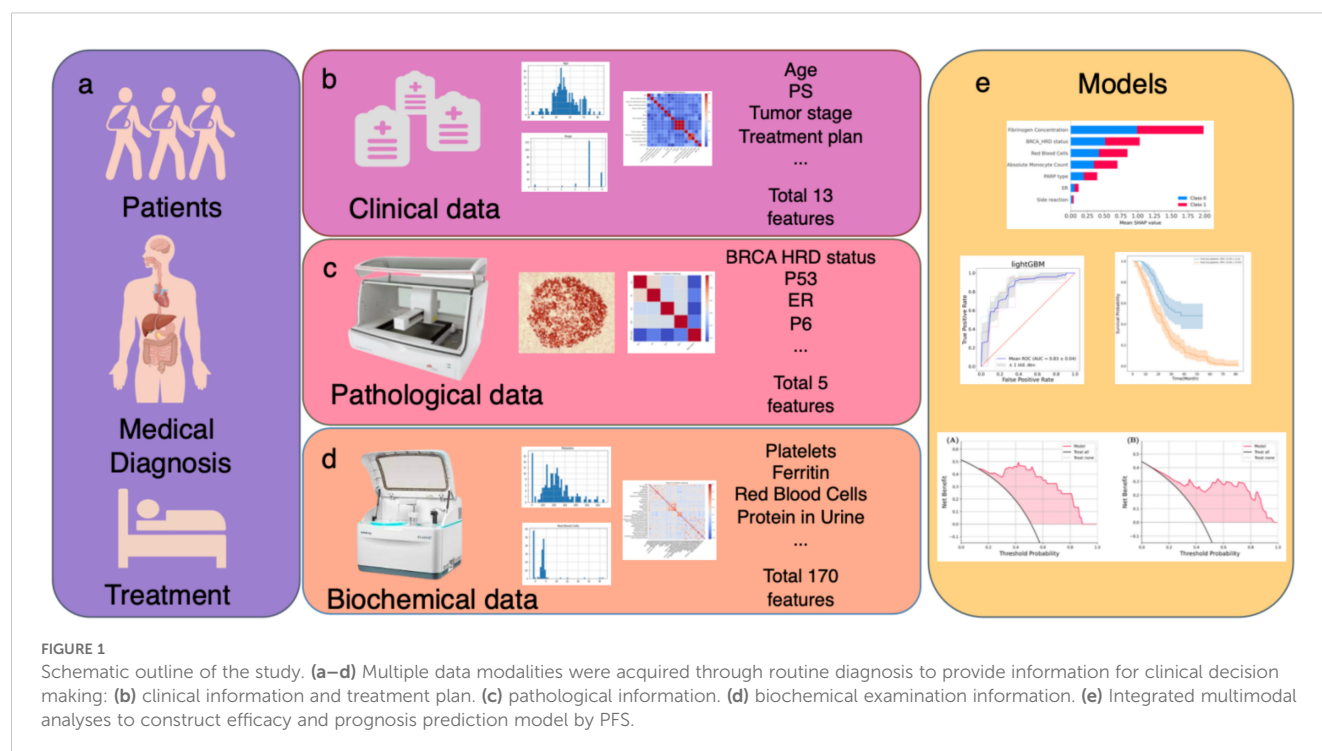
found that circulating tumor DNA was significantly associated with worse PFS and OS in patients with epithelial ovarian cancer (28). A recent review provides up-to-date evidence and summarizes the currently available therapeutic options for the treatment of ovarian cancer recurrence, investigating the factors that must be considered when choosing the best therapy, including molecular characterization and disease burden, while also presenting the limitations of current treatment options (29). Therefore, relying solely on BRCA mutation/HRD status as a clinical genetic indication for PARPi therapy is inadequate. Multiple factors—including clinical characteristics, tumor type and stage, quality of prior surgery, chemotherapy regimens, maintenance treatment protocols, biochemical markers, and pathological profiles—may influence PARPi efficacy and prognosis. Identifying key prognostic factors among a vast array of clinicopathological variables, eliminating redundancies, and constructing a robust and precise prediction model for PARPi efficacy and prognosis is complex. Traditional statistical approaches are often insufficient to address these complexities. Consequently, the integration of multimodal data using advanced machine learning techniques is essential. This approach promises to enhance the prediction of PARPi efficacy and refine prognostic assessments, thereby informing personalized treatment strategies.

Machine learning (ML) has been widely applied in the medical field recently (30, 31). Given the vast amount of medical data, intricate patterns, and individual-specific expressions, ML offers unique advantages. It can extract significant patterns from complex medical datasets and achieve optimal model performance by identifying the most contributory feature combinations. ML has been extensively employed in cancer prognosis research, providing effective and accurate prognostic conclusions based on cancer sample data (32–34). In this work, we set out to investigate the clinical, pathological, and biochemical information obtained during routine diagnostic and therapeutic work in patients with ovarian cancer (Figure 1a). We performed correlation analysis and feature screening for the three types of features (Figures 1b–d). Finally, ML algorithms were used to construct a prognostic prediction model (Figure 1e).

2 Methods

2.1 Patients

This multi-center retrospective study analyzed data from 251 ovarian cancer patients collected between August 2018 and November 2023. The cohort comprised 215 cases from the initial single-center dataset at Hunan Cancer Hospital, supplemented by 26 cases from Xiangya Hospital of Central South University and 10 cases from The Second Xiangya Hospital of Central South University through subsequent multicenter collaboration. The inclusion criteria were as follows: histologically confirmed ovarian cancer; R0 residual tumor status; received platinum-based chemotherapy; received PARPi as maintenance therapy for more than 3 months. Exclusion criteria were as follows: follow-up time



less than 6 months. This study was approved by the ethics committee of Hunan Cancer Hospital.

2.2 Data

Basic patient data were retrieved from the electronic medical record, and tumor recurrence and survival status were obtained by biweekly telephone follow-up. The dimensions of information collected included the patient's clinical data (age, gender, performance status, height, weight, body mass index, marriage age, number of pregnancies, number of abortions, history of endocrine chronic diseases, history of cardiovascular diseases, history of infectious diseases, history of other cancers, pathological stage, pathological classification, metastasis type and platinum sensitive status); immunohistochemical data (P53, ER, P16, Ki67); biochemical data (blood routine, urine routine et al.); treatment data (treatment line, chemotherapy, PARPi type, toxicity and side effects, primary surgery hospital and secondary surgery hospital); BRCA mutation/HRD status; outcome data (patient recurrence status). Among the outcome measures, we agreed that a good outcome was defined as a PFS ≥ 24 months in primary ovarian cancer patients and a PFS ≥ 12 months in recurrent ovarian cancer patients. The classification into the primary or recurrent ovarian cancer patient cohort was determined based on the timing of the patients' first use of PARPi. Patients were categorized into the primary cohort if their first use of PARPi occurred during the initial diagnosis of ovarian cancer, while those who first used PARPi during the recurrent phase were classified into the recurrent patient cohort.

2.3 Statistical analysis

In the exploration phase of data analysis, the Shapiro-Wilk test was used to test the normality of features, and the Spearman correlation coefficient was used to analyze the correlation between features. The Cox proportional hazards regression model is a statistical method used in survival analysis. We used the Cox proportional hazards model to perform univariate analysis for all variables. Variables that were significant with a P value of less than 0.05 in the Cox univariate analysis were entered into the subsequent Cox multivariate analysis. All the above analyses were implemented using Python3.6, where the Shapiro-Wilk test is from the Scipy library (35), the Spearman correlation coefficient analysis was performed using Pandas library (36), and Cox variable analysis was performed using Lifelines library (37).

2.4 Machine learning

2.4.1 Feature selection

To ensure the objectivity and stability of feature selection, the following features screening strategy was conducted for clinical, pathological, and biochemical features:

2.4.1.1 Remove features with a correlation value greater than 0.8

The correlation is calculated for every pair of features. For feature pairs with a correlation greater than 0.8, the occurrence frequency of each feature is counted, and the feature with the highest frequency is removed. The correlation among the remaining

features is then recalculated, and this process is repeated until the correlation between all features is less than 0.8.

2.4.1.2 Retain features with P values less than 0.05

Each feature was assessed by fitting a univariate Cox proportional hazards regression model using the Python Lifelines software package. The risk model was computed on the training set, and univariate coefficients and significance confidence were plotted. For features where the model failed to converge, the fit was retried using L2 regularization with a parameter $c = 0.2$. If the model still failed to converge, a P value of 1 and a hazard ratio (HR) of 1 were assigned. Features with P values less than 0.05 were retained for subsequent analysis.

2.4.1.3 Exclude features with a VIF greater than 10

VIF is a statistical tool used to detect multicollinearity among features, which can lead to model instability and diminish both explanatory and predictive performance. VIF measures the linear relationship between each feature and the others by performing linear regression on each feature against the others, calculating the regression coefficient, and determining the R^2 of the regression model. VIF is then calculated using the formula $VIF = \frac{1}{1-R^2}$. Generally, VIF values greater than 10 indicate strong collinearity among features. During feature screening, features with high VIF values are eliminated to reduce the impact of multicollinearity.

2.4.1.4 Preserve stable features

Elastic Net is a linear regression method that combines L1 regularization (Lasso) and L2 regularization (Ridge), which can be used for feature selection and model stability enhancement. When using the Elastic Net method for feature selection, the stability and reliability of selected features are assessed by observing changes in different data sets and model parameters. In this study, the scikit-learn library is used to implement feature selection for Elastic Net stability (38).

2.4.2 Model construction

In the process of model construction, 7 common ML algorithms are selected to build the model (38, 39).

- Linear model: Logistic Regression(LR)
- Nearest neighbor method: K-Nearest Neighbors (KNN)
- Ensemble method: Random Forest (RF), lightGBM, and XGBoost
- Support Vector Machine: Support Vector Machine (SVM)
- Probabilistic model: Naive Bayes

In the data preprocessing stage, features with more than 50% missing data were deleted. Given that KNN and SVM algorithms are highly sensitive to data scale, Z-score normalization was applied for these models, while normalization was deemed unnecessary for other models.

2.4.3 Model parameters tuning

The performance of a ML model is often influenced by its parameters, which are tuned to optimize the model's performance and enhance its generalization capability. The primary objective of parameter tuning is to identify the optimal combination of parameters that can maximize or minimize performance metrics (such as accuracy, precision, recall, etc.), while preventing overfitting and improving the model's ability to generalize. During the model tuning stage, the GridSearch method is employed to conduct a comprehensive parameter search (38).

2.4.4 Model evaluation

The internal cohort comprised 215 ovarian cancer cases from Hunan Cancer Hospital (August 2018–November 2023), which were randomly partitioned into training and internal validation sets at a 4:1 ratio for model development. External validation was conducted using two datasets: 26 cases from Xiangya Hospital of Central South University and 10 cases from The Second Xiangya Hospital of Central South University, collectively forming a 36-case multicenter validation cohort.

In the model training and validation process, five-fold cross-validation was utilized. The dataset was randomly divided into five equal subsets. Each time, four subsets were used for training, and the remaining subset was used for testing. This process was repeated five times, and the average of the results was calculated to reduce the bias caused by different data partitions. The performance indicators selected were AUC, accuracy, F1 score, sensitivity, and specificity.

To assess the generalization ability of a ML model, a key step is to ensure that the model performs well on unseen data (40). However, numerous factors influence this ability, including model complexity, training set size, consistency and stability of data distribution between training and test sets, and the distribution state of the loss function in parameter space. Prior research has extensively examined model generalization across various models and application domains (40–43). Commonly employed techniques for evaluating generalization encompass comparing model performance on training and test sets, cross-validation, and multi-index evaluation. Discrepancies between performance on training and test sets may indicate overfitting or underfitting. Additionally, comprehensive evaluation entails consideration of multiple performance metrics such as accuracy, precision, recall, F1 score, and area under the ROC curve (AUC). While AUC serves as a primary metric for generalization evaluation, offering a holistic view of model performance across classification thresholds and robustness to class imbalance, it lacks specificity in assessing individual category performance, particularly for rare events. Moreover, AUC's inability to directly address overfitting and underfitting limits its comprehensive assessment of model generalization.

To address the limitations of current generalization metrics, we propose the concept of Model Generalization Ability (MGA). MGA integrates the AUC metric on the test set (as the primary generalization metric) with the consistency of AUC between the

training and test sets to evaluate both the stability and generalization ability of various ML models on a given dataset. We define MGA as:

$$MGA = AUC_{\text{test set}} \times R_{\text{model generalization}}$$

where $AUC_{\text{test set}}$ is the AUC value of the model on the test set, and $R_{\text{model generalization}}$ is the ratio of the test set AUC to the training set AUC, given by:

$$R_{\text{model generalization}} = \frac{AUC_{\text{test set}}}{AUC_{\text{training set}}}$$

A higher $AUC_{\text{test set}}$ indicates that the model generalizes well to unseen data, which is often prioritized over training set performance when evaluating different models. This is because test set performance more accurately reflects the model's true generalization ability. Conversely, a higher $R_{\text{model generalization}}$ denotes greater consistency between the model's performance on the training and test sets, signifying enhanced model stability. A model exhibiting both high $AUC_{\text{test set}}$ and $R_{\text{model generalization}}$ demonstrates superior generalization power, characterized by stable outputs and robust predictive performance on unfamiliar data. This dual consideration not only meets the requirements for evaluating generalization ability but also mitigates the issues of multiple comparisons inherent in multi-metric evaluations. By incorporating both performance and consistency, MGA provides a comprehensive assessment of a model's generalization capabilities.

Finally, calibration curve analyses were utilized to compare the agreement between predicted probabilities and observed outcomes. Decision Curve Analysis (DCA) was performed to quantify the net benefits across different threshold probabilities, thereby assessing the clinical utility of the model and determining its effectiveness under various threshold probabilities.

2.4.5 Model interpretation

To account for model features interpretability, we use SHapley Additive exPlanations (SHAP), which provides a systematic and unbiased approach to interpreting the predictions of ML models. The advantages of SHAP are as follows:

- **Interpretability:** SHAP enhances the interpretability of model predictions by assigning weights to each feature. This enables users to understand the model's dependency on different features, thereby gaining more insight into the model's decision-making process.
- **Fairness:** Based on Shapley values, SHAP ensures a fair contribution of each feature. This method appropriately weights each feature, preventing excessive emphasis on any single feature and helping to avoid bias or unfairness in model interpretation.
- **Increase in trust:** Understanding model predictions through SHAP can increase trust in the model. When users comprehend the reasons behind specific predictions, they are more likely to trust the model's reliability, particularly in critical decision-making domains such as healthcare and finance.

- **Problem diagnosis:** SHAP facilitates the identification of model weaknesses and potential issues. By analyzing SHAP values, users can determine which features most significantly impact the model's predictions, contributing to improvements in the model's performance and robustness.

3 Results

3.1 Patients and disease characteristics

Table 1 presents the distribution of clinical and pathological characteristics for the internal cohort of 215 patients, respectively. The mean age of the patients was 55.40 ± 8.73 years. Regarding prior medical conditions, 3.26% of the patients had chronic endocrine-related diseases, 13.95% had cardiovascular-related diseases, 1.86% had infectious diseases, and 8.37% had combined other cancers. At initial presentation, 65.12% of the patients had stage III tumors, and 22.79% had stage IV tumors, with 91.16% exhibiting metastases at diagnosis. 40.93% of the patients were primary ovarian cancer patients, and 59.07% of the patients were recurrent ovarian cancer patients. In the primary ovarian cancer patients, not all patients experienced recurrence, with 47.7% of the patients having already relapsed. Among the recurrent ovarian cancer patients, patients with their first recurrence accounted for 73.2%, while those with second or subsequent recurrences accounted for 27.8%. Taxol plus platinum (TP) regimen was administered to 40.93% of the patients, and TP plus bevacizumab to 50.70%. For maintenance therapy, olaparib was used in 56.28% of the patients, niraparib in 31.02%, and both olaparib and niraparib in 2.79%.

3.2 Statistical analysis

The mean PFS was 27.93 ± 11.00 months for primary ovarian cancer patients and 23.04 ± 15.65 months for recurrent ovarian cancer patients (Supplementary Figure 1). Significant differences were observed in the PFS distribution between primary and recurrent ovarian cancer patients ($P < 0.001$). Consequently, statistical analyses and predictive model construction were conducted separately for these patient cohorts and Cox proportional hazards regression analysis was performed for all characteristics. The results for primary and recurrent ovarian cancer patients were presented in Table 2. Only variables with $P < 0.05$ in the univariate Cox analysis were included in the tables. Among primary ovarian cancer patients, the variables significantly associated with PFS ($P < 0.05$) in the univariate analysis included BRCA mutation/HRD status, absolute value of lymphocytes (AVOL), PARPi type, antibody-ABO, total bile acids (TBAs), fibrinogen concentration, and thrombin time. While in the multivariate Cox analysis, only TBAs remained significant ($P = 0.04$). The hazard ratios (HRs) from both univariate and

TABLE 1 Clinical characteristics of the patients (n=215). Categorical variables were expressed as Count(Percent), and continuous variables were expressed as Mean ± SD.

Characteristics	Statistics
Age	55.40 ± 8.73 years
Marriage age	23.19 ± 3.05 years
Height	1.56 ± 0.05 m
Weight	54.09 ± 8.03 kg
BMI	22.16 ± 3.08 kg/m ²
Number of pregnancies	
0	4 (1.86)
1	34 (15.81)
2	53 (24.65)
>=3	103 (47.91)
Missing	21 (9.77)
Number of births	
0	7 (3.26)
1	81 (37.67)
2	81 (37.67)
>=3	25 (11.63)
Missing	21 (9.77)
PS	
0 score	194 (90.23)
1 score	18 (8.37)
2 score	3 (1.40)
Pathological type	
Serous	179 (83.26)
Mucinous	30 (13.95)
Clear cell	2 (0.93)
Endometrioid	4 (1.86)
Stage	
I	4 (1.86)
II	12 (5.58)
III	140 (65.12)
IV	49 (22.79)
Missing	10 (4.65)
Tumor metastasis type	
Without metastasis	19 (8.84)
Organ metastasis	66 (30.70)
Abdominal cavity, uterus or intestine metastasis	130 (60.47)

(Continued)

TABLE 1 Continued

Characteristics	Statistics
Primary surgery hospital	
Cancer Hospital	114 (53.02)
Non Cancer Hospital	101 (46.98)
Chronic endocrine history	
No	208 (96.74)
Yes	7 (3.26)
History of cardiovascular disease	
No	185 (86.05)
Yes	30 (13.95)
History of infectious diseases	
No	211 (98.14)
Yes	4 (1.86)
History of other tumors	
No	197 (91.63)
Yes	18 (8.37)
Platinum sensitive	
Yes	204 (94.88)
No	11 (5.12)
Treatment before PARP	
TP	88 (40.93)
TP + Bevacizumab	109 (50.7)
Others	18 (8.37)
PARPi type	
Olaparib	121 (56.28)
Niraparib	71 (33.02)
Olaparib + Niraparib	6 (2.79)
Fluzoparib	16 (7.44)
Missing	1 (0.47)
Treatment line	
Primary ovarian cancer	88 (40.93)
Recurrent ovarian cancer	127 (59.07)
Second Surgery Hospital	
Hunan Cancer Hospital	90 (41.86)
Others	29 (13.49)
Missing	96 (44.65)
BRCA mutation/HRD status	
No	102 (47.44)
Yes	113 (52.56)

(Continued)

TABLE 1 Continued

Characteristics	Statistics
P53	
No	30 (13.95)
Yes	112 (52.09)
Missing	73 (33.95)
ER	
Weak	72 (33.49)
Medium	50 (23.26)
Strong	14 (6.51)
Missing	79 (36.74)
P16	
No	9 (4.19)
Yes	91 (42.33)
Missing	115 (53.49)
Ki67	51.01 ± 23.14%

multivariate analyses indicate that BRCA mutation/HRD status is a protective factor; BRCA mutation/HRD positivity reduces the risk of recurrence, consistent with the conclusion of numerous current studies (10–12). Conversely, TBAs was identified as a risk factor in the multivariate analysis, with each unit increase in TBAs associated with a 10% increase in recurrence risk. In the cohort of recurrent ovarian cancer patients, significant variables ($P < 0.05$) in the univariate analysis included Ki67, isoenzymes of aspartate aminotransferase (IOAA), fasting blood glucose (FBG), glycated hemoglobin, uric acid, and CA-199. In the multivariate Cox analysis, only CA-199 remained significant ($P < 0.01$).

3.3 Machine learning

We employed ML to construct PARPi response and prognosis prediction models in ovarian cancer, consisting of two main stages: feature selection and model construction/prediction. During the feature selection stage, correlation analysis was performed for all features. Features were screened based on the principle of excluding highly correlated features (Supplementary Figures 2, 3). Supplementary Tables 1, 2 present the features with $P < 0.05$

TABLE 2 Results of Cox univariate and multivariate analysis of primary(n=88) and recurrent(n=127) ovarian cancer patients respectively.

Patient category	Feature	Univariate		Multivariate	
		HR	P	HR	P
Primary ovarian cancer patients	PARPi type	1.38 (1.02, 1.86)	0.04	1.3 0.92, 1.83)	0.13
	AVOL ^a	0.74 (0.56, 1.0)	0.05	0.87 (0.63, 1.2)	0.39
	Antibody-ABO ^b	1.0 (1.0, 1.0)	0.02	1.0 (1.0, 1.0)	0.07
	TBAs ^c	1.09 (1.02, 1.18)	0.02	1.10 (1.0, 1.2)	0.04
	Thrombin time ^d	1.05 (1.0, 1.09)	0.03	0.98 (0.91, 1.06)	0.69
	Fibrinogen concentration ^e	1.16 (1.05, 1.29)	<0.01	1.16(0.98, 1.38)	0.08
	BRCA mutation/HRD status	0.49 (0.26, 0.89)	0.02	0.75 (0.38, 1.50)	0.42
Recurrent ovarian cancer patients	Ki67	1.36 (1.05, 1.75)	0.02	1.27 (0.98, 1.65)	0.07
	IOAA ^f	1.01 (1.0, 1.02)	<0.01	1.01 (1.0, 1.02)	0.07
	FBG ^g	1.13 (1.04, 1.23)	<0.01	1.05 (0.95, 1.16)	0.37
	Glycated Hemoglobin ^h	1.12 (1.05, 1.20)	<0.01	1.07 (0.99, 1.15)	0.11
	Uric Acid ⁱ	1.0 (1.0, 1.0)	<0.01	1.0 (1.0, 1.0)	0.16
	CA-199 ^j	1.0 (1.0, 1.01)	<0.01	1.0 (1.0, 1.01)	<0.01

^aBlood routine: Absolute value of lymphocytes.

^bABO blood group: specific antibody-ABO blood type.

^cRoutine liver function tests: Total Bile Acids.

^dBlood clotting routine four items: Thrombin time.

^eBlood clotting routine four items: fibrinogen concentration.

^fRoutine liver function tests: Isoenzymes of Aspartate Aminotransferase.

^gFasting blood glucose: Fasting blood glucose.

^hDetermination of hemoglobin A1c: Glycated Hemoglobin.

ⁱRoutine renal function program: Uric Acid.

^jCarbohydrate antigen-199 (CA-199): Carbohydrate antigen-199.

following Cox univariate analysis. Subsequently, VIF values and elastic net feature stability values were calculated. Features with VIF ≥ 10 and stability ≤ 0.7 were excluded. For primary ovarian cancer patients, the final model included the following features: BRCA mutation/HRD status, PARPi type, antibody-ABO, TBAs, fibrinogen concentration, and thrombin time. For recurrent ovarian cancer patients, the final model incorporated Ki67, IOAA, FBG, glycated hemoglobin, uric acid, and CA-199.

We selected seven ML models—LR, KNN, RF, LightGBM, XGBoost, SVM, and Naive Bayes—to construct prediction models for both primary and recurrent ovarian cancer patient data. Optimal parameter combinations for each model were identified through grid search, and model evaluation was conducted using 5-fold cross-validation. Figure 2 illustrates the ROC curves for each of the seven models on internal and external datasets for primary and recurrent ovarian cancer patients. Table 3 presents a comparison of

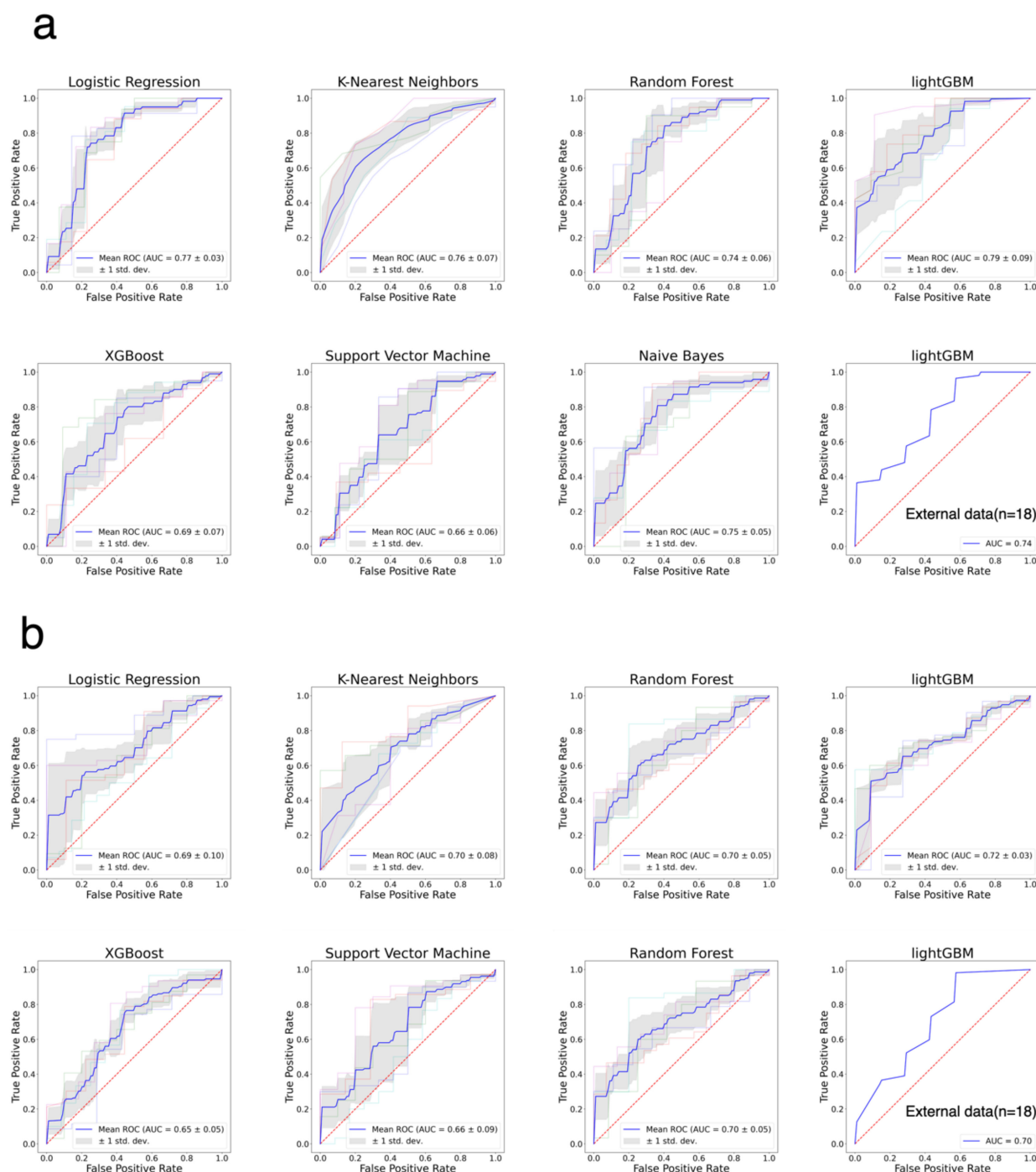


FIGURE 2

The ROC curves for the classification of patients with good and bad PFS patients. (a) ROC curves of seven models for primary ovarian cancer patients. (b) ROC curves of seven models for recurrent ovarian cancer patients.

TABLE 3 Summary of ML algorithms predictive performance for the primary and recurrent ovarian cancer patients(Mean $\pm$ SD).

Patient category	Dataset	models	AUC	Accuracy	Sensitivity	Specificity	F1 score	Precision	Recall	MGA
Primary ovarian cancer patients	Internal	LR	0.77 $\pm$ 0.03	0.69 $\pm$ 0.07	0.73 $\pm$ 0.15	0.61 $\pm$ 0.18	0.74 $\pm$ 0.08	0.78 $\pm$ 0.06	0.73 $\pm$ 0.15	0.77 $\pm$ 0.09
		KNN	0.76 $\pm$ 0.07	0.62 $\pm$ 0.08	0.54 $\pm$ 0.16	0.73 $\pm$ 0.08	0.61 $\pm$ 0.14	0.72 $\pm$ 0.09	0.54 $\pm$ 0.16	0.71 $\pm$ 0.14
		RF	0.74 $\pm$ 0.06	0.74 $\pm$ 0.03	0.89 $\pm$ 0.03	0.42 $\pm$ 0.07	0.82 $\pm$ 0.02	0.76 $\pm$ 0.04	0.89 $\pm$ 0.03	0.60 $\pm$ 0.11
		SVM	0.66 $\pm$ 0.06	0.61 $\pm$ 0.06	0.6 $\pm$ 0.18	0.65 $\pm$ 0.21	0.65 $\pm$ 0.1	0.78 $\pm$ 0.08	0.6 $\pm$ 0.18	0.44 $\pm$ 0.08
		lightGBM	0.79 $\pm$ 0.09	0.61 $\pm$ 0.1	0.65 $\pm$ 0.19	0.63 $\pm$ 0.27	0.68 $\pm$ 0.1	0.78 $\pm$ 0.16	0.65 $\pm$ 0.19	0.77 $\pm$ 0.19
		XGBoost	0.69 $\pm$ 0.07	0.63 $\pm$ 0.09	0.71 $\pm$ 0.24	0.57 $\pm$ 0.22	0.67 $\pm$ 0.15	0.74 $\pm$ 0.14	0.71 $\pm$ 0.24	0.48 $\pm$ 0.09
		Naive Bayes	0.75 $\pm$ 0.05	0.75 $\pm$ 0.03	0.77 $\pm$ 0.12	0.67 $\pm$ 0.19	0.78 $\pm$ 0.06	0.81 $\pm$ 0.04	0.77 $\pm$ 0.12	0.72 $\pm$ 0.12
	External	lightGBM	0.74	0.62	0.62	0.63	0.61	0.77	0.62	–
Recurrent ovarian cancer patients	Internal	LR	0.69 $\pm$ 0.1	0.68 $\pm$ 0.15	0.73 $\pm$ 0.26	0.54 $\pm$ 0.27	0.75 $\pm$ 0.18	0.86 $\pm$ 0.08	0.73 $\pm$ 0.26	0.68 $\pm$ 0.24
		KNN	0.7 $\pm$ 0.08	0.6 $\pm$ 0.07	0.59 $\pm$ 0.11	0.69 $\pm$ 0.18	0.69 $\pm$ 0.07	0.86 $\pm$ 0.08	0.59 $\pm$ 0.11	0.49 $\pm$ 0.11
		RF	0.7 $\pm$ 0.05	0.55 $\pm$ 0.08	0.42 $\pm$ 0.16	0.86 $\pm$ 0.13	0.55 $\pm$ 0.16	0.88 $\pm$ 0.08	0.42 $\pm$ 0.16	0.49 $\pm$ 0.07
		SVM	0.66 $\pm$ 0.09	0.68 $\pm$ 0.08	0.71 $\pm$ 0.15	0.62 $\pm$ 0.19	0.76 $\pm$ 0.08	0.84 $\pm$ 0.08	0.71 $\pm$ 0.15	0.46 $\pm$ 0.13
		lightGBM	0.72 $\pm$ 0.03	0.6 $\pm$ 0.11	0.55 $\pm$ 0.22	0.74 $\pm$ 0.28	0.65 $\pm$ 0.14	0.9 $\pm$ 0.07	0.55 $\pm$ 0.22	0.76 $\pm$ 0.05
		XGBoost	0.65 $\pm$ 0.05	0.57 $\pm$ 0.07	0.46 $\pm$ 0.13	0.84 $\pm$ 0.06	0.6 $\pm$ 0.12	0.88 $\pm$ 0.05	0.46 $\pm$ 0.13	0.43 $\pm$ 0.06
		Naive Bayes	0.73 $\pm$ 0.08	0.63 $\pm$ 0.1	0.62 $\pm$ 0.18	0.65 $\pm$ 0.19	0.7 $\pm$ 0.13	0.86 $\pm$ 0.06	0.62 $\pm$ 0.18	0.75 $\pm$ 0.15
	External	lightGBM	0.70	0.61	0.50	0.67	0.71	0.82	0.60	–

various performance metrics for these models on primary and recurrent ovarian cancer patients, respectively. For primary ovarian cancer patients, the best-performing model according to the MGA was LightGBM, with an MGA of 0.77 ± 0.19 . The performance metrics for LightGBM in internal test set were: AUC = 0.79 ± 0.09 , accuracy = 0.61 ± 0.10 , sensitivity = 0.65 ± 0.19 , specificity = 0.63 ± 0.27 , F1 score = 0.68 ± 0.10 , precision = 0.78 ± 0.16 , and recall = 0.65 ± 0.19 . For recurrent ovarian cancer patients, LightGBM also demonstrated the best performance, with an MGA of 0.76 ± 0.05 . The performance metrics for this model in internal test set were: AUC = 0.72 ± 0.03 , accuracy = 0.60 ± 0.11 , sensitivity = 0.55 ± 0.22 , specificity = 0.74 ± 0.28 , F1 score = 0.65 ± 0.14 , precision = 0.90 ± 0.07 , and recall = 0.55 ± 0.22 . Furthermore, rigorous external validation conducted in independent multicenter cohorts confirmed the robustness of the LightGBM model, demonstrating discriminative performance with AUC values of 0.74 in the primary cohort and 0.70 in the recurrent cohort.

Complete metrics including sensitivity, specificity, and clinical utility metrics are systematically documented in Table 3.

The calibration curves for both the primary and recurrent ovarian cancer patient models are presented in Figures 3a, b. These curves evaluate the reliability of probability predictions by comparing the model-predicted probabilities (x-axis) with the observed positive rates (y-axis). The analysis reveals that the solid blue line (model predictions) closely approximates the dashed line (perfect calibration) across most intervals (e.g., 0.1–0.5), indicating accurate efficacy predictions for low-to-moderate probability ranges. However, in the high-probability interval (>0.6), the observed positive rates are slightly higher than the predicted probabilities, suggesting a potential mild underestimation of efficacy for high-probability samples. This finding underscores the importance of avoiding overreliance on a single high-probability threshold for clinical decision-making. The DCA curves for the primary and recurrent ovarian cancer patient models are illustrated

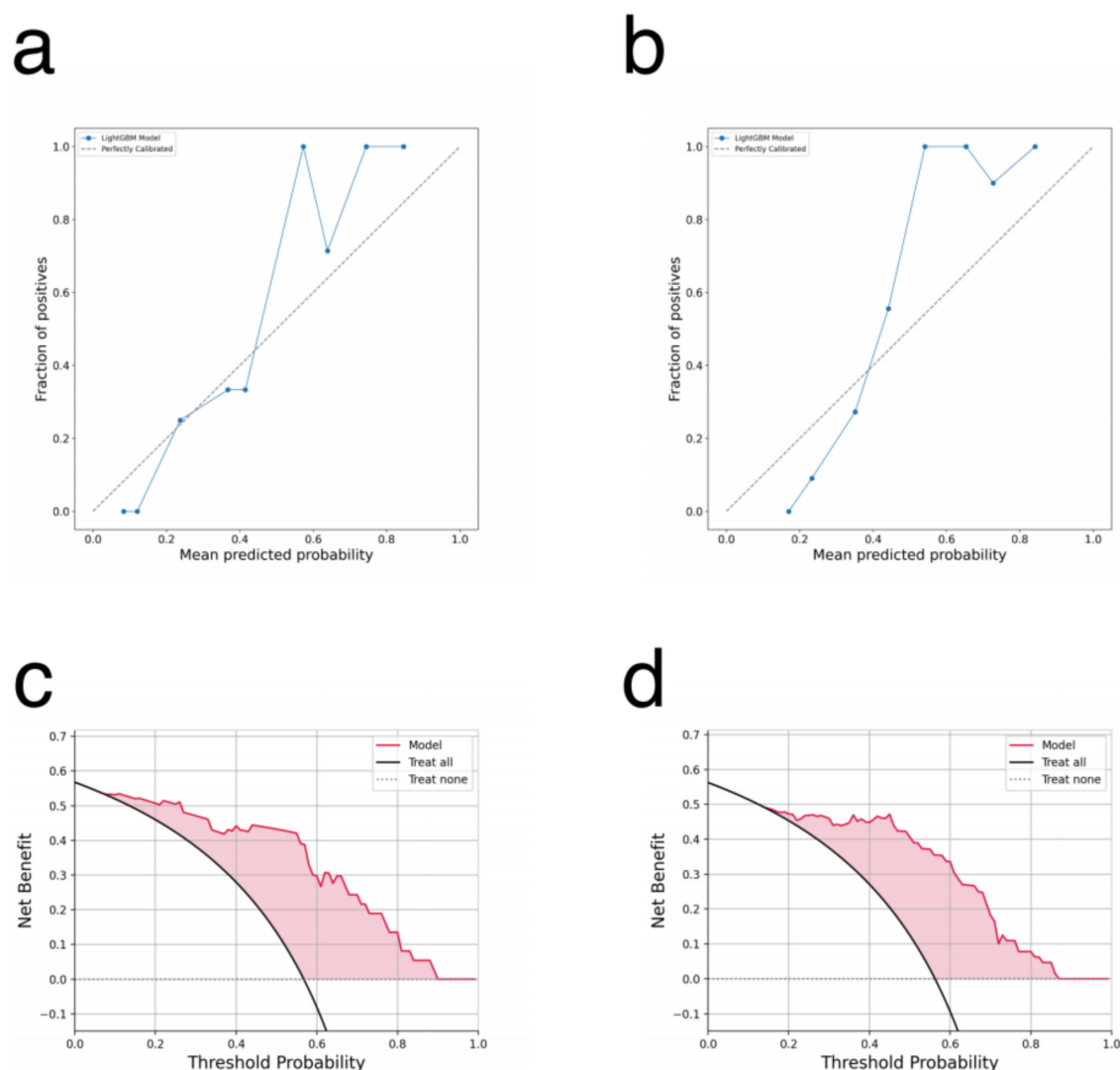


FIGURE 3

(a, b) The calibration curve for the lightGBM model: (a) primary ovarian cancer patients. (b) recurrent ovarian cancer patients. (c, d) The DCA curve for the lightGBM model: (c) primary ovarian cancer patients. (d) recurrent ovarian cancer patients.

in Figures 3c, d. These curves assess the clinical utility of the models across various decision-making scenarios by analyzing the relationship between threshold probabilities and net benefit. The red curve (model) demonstrates significantly higher net benefit than the black line (“Treat all”) and gray line (“Treat none”) across most threshold probability ranges. This indicates that employing the model to guide PARP inhibitor treatment could substantially enhance clinical net benefit within these intervals.

Figures 4a–d presents the results of interpreting the LightGBM model for primary ovarian cancer patients using SHAP. Figures 4a, b are global bar and scatter plots, respectively. Figure 4a illustrates the importance and rank of each feature’s contribution to the model, with BRCA mutation/HRD status, TBAs, and fibrinogen concentration ranking as the top three most important features. Figure 4b shows how each sample contributes to the model based on each feature. As depicted, samples with positive BRCA mutation/HRD status (value 1, represented by red dots) generally contribute positively to the model,

indicating a tendency towards lower relapse risk. In contrast, samples with negative BRCA mutation/HRD status (value 0, represented by blue dots) typically contribute negatively, indicating a higher tendency towards relapse. Additionally, higher values of TBAs and fibrinogen concentration contributed negatively to the model. Patients treated with olaparib (value 0 for PARPi type) generally contributed positively to the model, while other PARPi types contributed negatively. Figures 4c, d depict the contribution of each feature to the model for a single sample. For this particular sample, a BRCA mutation/HRD status of 1 provided a positive contribution of 0.74 to the model. A TBAs value of 2.8 $\mu\text{mol/L}$ (within the normal reference range of 0.5–10 $\mu\text{mol/L}$) contributed positively with a value of 0.03. Conversely, a fibrinogen concentration of 6.33 g/L (exceeding the normal reference range of 1.8–3.5 g/L) contributed negatively with a value of 0.28. These contributions are consistent with the overall trends observed in the global analysis.

Figures 4e–h presents the results of interpreting the LightGBM model using SHAP for recurrent ovarian cancer patients. Figures 4e, f

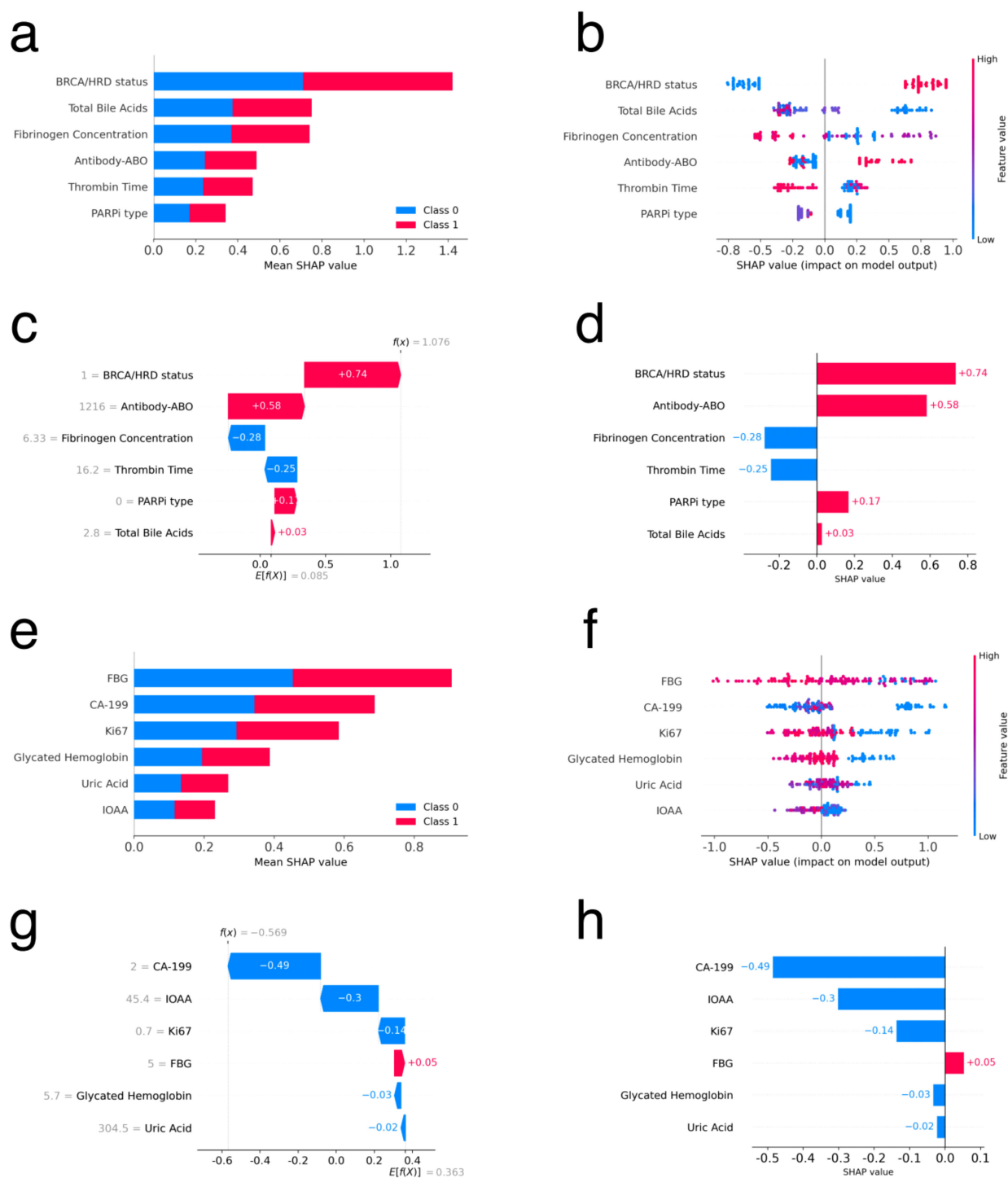


FIGURE 4

Feature interpretation using SHAP for primary and recurrent ovarian cancer patients. (a, e) Bar plot of feature importance sorted by mean SHAP value for primary and recurrent ovarian cancer patients respectively. (b, f) Density scatter plot of SHAP values for each feature for primary and recurrent ovarian cancer patients respectively. (c, g) Local waterfall plot of feature importance sorted by mean SHAP value for one sample for primary and recurrent ovarian cancer patients respectively. (d, h) Local bar plot of feature importance sorted by mean SHAP value for one sample for primary and recurrent ovarian cancer patients respectively.

are global bar and scatter plots, respectively. Figure 4e shows the importance and rank of each feature's contribution to the model, with the top three features being FBG, CA-199, and Ki67. Figure 4f illustrates the contributions of all samples to the model based on

each feature. The figure indicates a complex contribution trend, with no clear boundary to distinguish the impact of sample values on the model. However, in general, higher values of FBG, CA-199, Ki67, and glycated hemoglobin exhibit a negative trend in their contributions to

Figure 5 shows a web-based interactive interface titled "Efficacy and prognosis prediction of PARP inhibitors". It contains several input fields with pre-filled values: BRCA/HRD status (1), Total Bile Acids (2.8), Fibrinogen Concentration (6.33), Antibody-ABO (1216), Thrombin Time (16.2), and PARPi (0). Below these fields is a "Predict" button. At the bottom of the interface, it displays the result: "Patient's risk level: Low".

FIGURE 5
Interactive interface of the PARP inhibitor efficacy and prognosis prediction model for clinical use. Taking the PARP inhibitor efficacy and prognosis prediction model of primary ovarian cancer patient as an example, further details can be found at the GitHub link.

the model. Figures 4g, h depict the contribution of each feature to the model for a single sample. These figures help in understanding how individual features influence the model's predictions for specific cases.

4 Discussion

Ovarian cancer is widely recognized for its significant genomic disruption and high mutation rate (44). Clinical trials have demonstrated that patients with BRCA mutation or HRD can significantly benefit from PARPi (10–12). Interestingly, some non-HRD or BRCA-wild type patients may also experience benefits from PARPi (14–16). In the era of precision medicine, it is crucial to identify patients who will benefit from PARPi therapy, as from preoperative evaluation to postoperative treatment plans, personalized approaches for patients are receiving increasing attention (45, 46). At present, the prognostic factors influencing the efficacy of PARPi in ovarian cancer remain unclear. While clinical examinations, pathological and biochemical tests are readily available in practice, their combined predictive power is yet to be fully determined. In this study, we explored the impact of 188 potential influencing factors on PFS, including clinical, pathological, and biochemical information. Although classical statistical methods can identify associations between variables and outcomes (47), ML methods are more suitable for multivariate predictive classification tasks.

In this study, we collected data from 251 patients, encompassing clinical characteristics, pathological, and biochemical information, resulting in a total of 188 features. Among primary ovarian cancer patients, TBAs was identified as a significant prognostic factor for PFS through both Cox univariate ($P = 0.02$) and multivariate ($P = 0.04$) analyses. In the feature importance analysis of the ML model, the top three factors were BRCA mutation/HRD status, TBAs, and fibrinogen concentration. BRCA mutation/HRD status notably influences the prognosis of ovarian cancer. Consistent with many current studies,

BRCA mutation/HRD-positive patients in our study tended to have better prognoses (10–12, 14, 15). TBAs, which are cholesterol-derived sterols and signaling molecules, play a key role in regulating cancer cell behavior through receptor-mediated functions. The activation of receptors such as liver X receptor, pregnane X receptor, vitamin D receptor, or constitutive androstane receptor has been shown to protect against ovarian cancer (21–23, 25). These protective effects, similar to those elicited by bile acids (BAs), include the inhibition of proliferation, invasion, epithelial-mesenchymal transition, *de novo* fatty acid biosynthesis, and the proportion of the cancer stem cell population, as well as improving the efficacy of chemotherapy (48–51). Additionally, BAs can influence the expression and activity of various PARP enzymes (24), and deoxycholic acid can regulate the expression of BRCA1 and estrogen receptors, thereby controlling the drug sensitivity of ovarian cancer cells (52). Thus, BAs may modulate the efficacy of PARPi.

CA-199 was identified as a significant prognostic factor for PFS in recurrent ovarian cancer through both Cox univariate ($P < 0.01$) and multivariate ($P < 0.01$) analyses. In the feature importance analysis of the ML model, the top three factors were FBG, CA-199, and Ki67. Our study findings from both Cox factor analysis and the ML model suggested that CA-199 and FBG were risk factors affecting prognosis, with increases in CA-199 and FBG levels associated with an elevated risk of poor prognosis. CA-199 is associated with primary cancers of the gastrointestinal system and ovary, serving as a diagnostic marker for gastrointestinal and ovarian mucoid cancers (53, 54). Zhu et al. conducted a study analyzing serum CA-199 levels in patients with normal postoperative serum CA-125 levels, confirming that an increase in postoperative serum CA-199 level was an independent risk factor for PFS and OS. Patients with elevated serum CA-199 levels exhibited significantly lower 5-year PFS and OS rates compared to those with normal levels (27). Furthermore, postoperative serum CA-199 levels have been shown to help identify subgroups of patients with normal postoperative CA-125 levels who are at higher risk of recurrence and death.

Additionally, the FBG levels of patients may serve as an important prognostic indicator. In ovarian tumors, increased expression of the transmembrane protein glucose transporter 1, responsible for glucose uptake, has been associated with shortened survival in ovarian cancer patients. Lamkin et al. found in univariate analysis that higher blood glucose levels were associated with shorter survival times (hazard ratio = 1.88; $P < 0.05$). Multivariate analysis adjusted for staging revealed that higher blood glucose levels were associated with shorter survival times (hazard ratio = 2.01; $P = 0.04$) and disease-free interval (hazard ratio = 2.32; $P < 0.05$). These findings suggest the prognostic value of blood glucose levels in ovarian cancer (26). In 2019, elevated fasting blood glucose was identified as the highest risk factor for ovarian cancer deaths, with a global increase in age-standardized mortality rates attributed to elevated fasting blood glucose across all Socio-demographic Index quintiles. This trend may be related to diabetes comorbidity-related deaths (55). Currently, diabetes has been confirmed as an independent risk factor for ovarian cancer mortality (56), and elevated blood glucose levels in ovarian cancer

patients are predictive factors for poorer survival rates (26). Rapidly proliferating cancer cells benefit from the nutrient-rich microenvironment of high blood glucose levels, as they have increased metabolic demands and utilize glucose faster than healthy cells (57). Chronic hyperglycemia can increase oxidative stress, reduce the tumor-suppressive activity of adenosine monophosphate-activated protein kinase, impair the diversion capacity of hexosamine monophosphate, and lead to the accumulation of advanced glycation end products, affecting the mitogen-activated protein kinase NF-KB pathway (58). Angiogenesis is crucial for the growth of ovarian cancer, as solid tumors typically require the formation of new blood vessels to grow larger than 1-2mm (59). The expression of vascular endothelial growth factor(VEGF) has been found to be correlated with blood glucose levels, and VEGF is now recognized as an effective pro-angiogenic factor, suggesting that patients with uncontrolled diabetes may be more susceptible to ovarian cancer.

In this study, we employed ML techniques to construct separate efficacy prognosis prediction models for primary and recurrent ovarian cancer patients based on multimodal data. Both models demonstrated favorable performance, with AUC values of 0.79 ± 0.09 and 0.72 ± 0.03 , respectively. Furthermore, we identified the top three factors with the greatest impact on the models: for primary ovarian cancer patients, these were BRCA mutation/HRD status, TBAs, and fibrinogen concentration, while for recurrent ovarian cancer patients, they were FBG, CA-199, and Ki67. Considering the interpretability limitations of ML models, we conducted feature analysis using SHAP. In primary ovarian cancer patients, BRCA mutation/HRD status positivity contributed positively to the model, while higher values of TBAs and fibrinogen concentration had a negative impact on the model. Patients treated with olaparib showed a general positive contribution to the model, whereas other PARPi types exhibited a negative contribution. In recurrent ovarian cancer patients, higher values of FBG, CA-199, Ki67, and glycated hemoglobin trended towards negative contributions to the model overall. Machine learning models often encounter challenges in feature interpretation and application. To address this, we have developed a user-friendly interface tool, as shown in Figure 5. This tool assists clinicians in inputting the necessary variables for the model and provides the patient's risk level, offering guidance for treatment strategy. The tool can be easily deployed and run locally, greatly simplifying its use in clinical settings. For detailed code, deployment, and application instructions, please refer to the GitHub link: https://github.com/xiongsa/PARP_efficiency_prediction.

It should be noted that we intentionally explored data generated during standard treatment. Using these data instead of data specifically collected for computational modeling significantly reduces the adoption cost of the final model in clinical workflows, but these data were not collected specifically for modeling purposes. For most patients, we included BRCA gene mutation status but not HRD status, as complete HRD gene sequencing data was not available for all cases. While BRCA mutations are the first and most widely used genotype prognostic factor for PARPi efficacy in ovarian cancer, they are not sufficient to predict the efficacy of PARPi. Current research findings suggest that HRD-positive status

is an important prognostic factor for PARPi. Based on synthetic lethality mechanisms, HRD is more widely distributed in ovarian cancer than BRCA mutations. This is because HRD can be caused not only by deleterious BRCA mutations but also by genomic alterations or epigenetic inactivation of BRCA genes and other defects independent of BRCA (60, 61), and it is associated with the efficacy of PARPi (62, 63). Therefore, it is hoped that more complete HRD gene sequencing data can be obtained in the future.

These results may provide research directions for exploring effective and accurate prognostic factors for PARPi efficacy in ovarian cancer. Therefore, there is an urgent need for large-scale, prospective clinical studies to explore effective and accurate prognostic factors for PARPi efficacy, thereby facilitating personalized PARPi treatment and expanding the use of PARPi to a more suitable population of ovarian cancer patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Hunan Cancer Hospital Ethical Review Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

XX: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. LC: Data curation, Methodology, Writing – original draft. NW: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. QN: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. ZY: Data curation, Methodology, Validation, Writing – review & editing. ZC: Data curation, Methodology, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by grants from the Hunan Provincial Natural Science Foundation of China(No.: 2023JJ30373 and 2023JJ30375), the National Key R&D

Program of China (No.: 2022YFC2404604), the Science and Technology Innovation Program of Hunan Province (No.: 2023SK4034), the Hunan Cancer Hospital Climb Plan (No.: 2023NSFC-A004 and 2023NSFC-A003).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2024) 74:229–63. doi: 10.3322/caac.21834
- Ledermann JA. PARP inhibitors in ovarian cancer. *Ann Oncol.* (2016) 27:140–4. doi: 10.1093/annonc/mdw094
- Cooke SL, Brenton JD. Evolution of platinum resistance in high-grade serous ovarian cancer. *Lancet Oncol.* (2011) 12:1169–74. doi: 10.1016/S1470-2045(11)70123-1
- McMeekin DS, Tillmanns T, Chaudry T, Gold M, Johnson G, Walker J, et al. Timing isn't everything: an analysis of when to start salvage chemotherapy in ovarian cancer. *Gynecol Oncol.* (2004) 95:157–64. doi: 10.1016/j.ygyno.2004.07.008
- Pfisterer J, Ledermann JA. Management of platinum-sensitive recurrent ovarian cancer. *Semin Oncol.* (2006) 33:S12–6. doi: 10.1053/j.seminoncol.2006.03.012
- Ashworth A. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol.* (2008) 26:3785–90. doi: 10.1200/JCO.2008.16.0812
- Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. *Science.* (2017) 355:1152–8. doi: 10.1126/science.aam7344
- O'Connor MJ. Targeting the DNA damage response in cancer. *Mol Cell.* (2015) 60:547–60. doi: 10.1016/j.molcel.2015.10.040
- Ray-Coquard I, Leary A, Pignata S, Cropet C, González-Martín A, Marth C, et al. Olaparib plus bevacizumab first-line maintenance in ovarian cancer: final overall survival results from the PAOLA-1/ENGOT-ov25 trial. *Ann Oncol.* (2023) 34:681–92. doi: 10.1016/j.annonc.2023.05.005
- Swisher EM, Lin KK, Oza AM, Scott CL, Giordano H, Sun J, et al. Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *Lancet Oncol.* (2017) 18:75–87. doi: 10.1016/S1470-2045(16)30559-9
- Moore K, Colombo N, Scambia G, Kim BG, Oaknin A, Friedlander M, et al. Maintenance olaparib in patients with newly diagnosed advanced ovarian cancer. *N Engl J Med.* (2018) 379:2495–505. doi: 10.1056/NEJMoa1810858
- Pujade-Lauraine E, Ledermann JA, Selle F, GebSKI V, Penson RT, Oza AM, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Oncol.* (2017) 18:1274–84. doi: 10.1016/S1470-2045(17)30469-2
- Arcieri M, Tius V, Andreetta C, Restaino S, Biasioli A, Poletto E, et al. How BRCA and homologous recombination deficiency change therapeutic strategies in ovarian cancer: a review of literature. *Front Oncol.* (2024) 14:1335196. doi: 10.3389/fonc.2024.1335196
- Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. *Lancet Oncol.* (2014) 15:852–61. doi: 10.1016/S1470-2045(15)70153-1
- Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N Engl J Med.* (2012) 366:1382–92. doi: 10.1056/NEJMoa1105535
- Mirza MR, Monk BJ, Herrstedt J, Oza AM, Mahner S, Redondo A, et al. Niraparib maintenance therapy in platinum-sensitive, recurrent ovarian cancer. *N Engl J Med.* (2016) 375:2154–64. doi: 10.1056/NEJMoa1611310
- Fong PC, Yap TA, Boss DS, Carden CP, Mergui-Roelvink M, Gourley C, et al. Poly(ADP-ribose) polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. *J Clin Oncol.* (2010) 28:2512–9. doi: 10.1200/JCO.2009.26.9589
- Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-McGuinn KM, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet.* (2010) 376:245–51. doi: 10.1016/S0140-6736(10)60893-8
- Huang XZ, Jia H, Xiao Q, Li RZ, Wang XS, Yin HY, et al. Efficacy and prognostic factors for PARP inhibitors in patients with ovarian cancer. *Front Oncol.* (2020) 10:958. doi: 10.3389/fonc.2020.00958
- Kobayashi Y, Banno K, Aoki D. Current status and future directions of ovarian cancer prognostic models. *J Gynecol Oncol.* (2021) 32:e34. doi: 10.3802/jgo.2021.32.e34
- Curtarello M, Tognon M, Venturoli C, Silic-Benussi M, Grassi A, Verza M, et al. Rewiring of lipid metabolism and storage in ovarian cancer cells after anti-VEGF therapy. *Cells.* (2019) 8:1601. doi: 10.3390/cells8121601

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2025.1571193/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

The PFS survival curve of primary and recurrent ovarian cancer patients. The blue and yellow line with the error bar shows the results of primary and recurrent ovarian cancer respectively.

SUPPLEMENTARY FIGURE 2

The Spearman correlation of primary ovarian cancer patients. (a) Spearman correlation between patient clinical characteristics; (b) Spearman correlation between pathological features of patients; (c) Spearman correlation between patient biochemical omics features; All three figures are symmetric along the diagonal.

SUPPLEMENTARY FIGURE 3

The Spearman correlation of recurrent ovarian cancer patients. (a) Spearman correlation between patient clinical characteristics; (b) Spearman correlation between pathological features of patients; (c) Spearman correlation between patient biochemical omics features; All three figures are symmetric along the diagonal.

SUPPLEMENTARY TABLE 1

Characteristic VIF and stability results in groups of primary ovarian cancer patients.

SUPPLEMENTARY TABLE 2

Characteristic VIF and stability results in groups of recurrent ovarian cancer patients.

22. Chen Y, Tang Y, Guo C, Wang J, Boral D, Nie D. Nuclear receptors in the multidrug resistance through the regulation of drug-metabolizing enzymes and drug transporters. *Biochem Pharmacol.* (2012) 83:1112–26. doi: 10.1016/j.bcp.2012.01.012
23. Wang Y, Masuyama H, Nobumoto E, Zhang G, Hiramatsu Y. The inhibition of constitutive androstane receptor-mediated pathway enhances the effects of anticancer agents in ovarian cancer cells. *Biochem Pharmacol.* (2014) 90:356–66. doi: 10.1016/j.bcp.2014.05.017
24. Szántó M, Gupta R, Kraus WL, Pacher P, Bai P. PARPs in lipid metabolism and related diseases. *Prog Lipid Res.* (2021) 84:101117. doi: 10.1016/j.plipres.2021.101117
25. Silvagno F, Poma CB, Realmuto C, Ravarino N, Ramella A, Santoro N, et al. Analysis of vitamin D receptor expression and clinical correlations in patients with ovarian cancer. *Gynecol Oncol.* (2010) 119:121–4. doi: 10.1016/j.ygyno.2010.04.010
26. Lamkin DM, Spitz DR, Shahzad MM, Zimmerman B, Lenihan DJ, Degeest K, et al. Glucose as a prognostic factor in ovarian carcinoma. *Cancer.* (2009) 115:1021–7. doi: 10.1002/cncr.24094
27. Zhu J, Jiang L, Wen H, Bi R, Wu X, Ju X. Prognostic value of serum CA19–9 and perioperative CA-125 levels in ovarian clear cell carcinoma. *Int J Gynecol Cancer.* (2018) 28:1108–16. doi: 10.1097/IGC.0000000000001293
28. Taliento C, Morciano G, Nero C, Froyman W, Vizzielli G, Pavone M, et al. Circulating tumor DNA as a biomarker for predicting progression-free survival and overall survival in patients with epithelial ovarian cancer: a systematic review and meta-analysis. *Int J Gynecol Cancer.* (2024) 34:906–18. doi: 10.1136/ijgc-2024-005313
29. Arcieri M, Andreetta C, Tius V, Zapelloni G, Titone F, Restaino S, et al. Molecular biology as a driver in therapeutic choices for ovarian cancer. *Int J Gynecol Cancer.* (2024). doi: 10.1136/ijgc-2024-005700
30. Benzekry S. Artificial intelligence and mechanistic modeling for clinical decision making in oncology. *Clin Pharmacol Ther.* (2020) 108:471–86. doi: 10.1002/cpt.1951
31. Breiman L. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Stat Science.* (2001) 16:199–231. doi: 10.1214/ss/1009213726
32. Zhu W, Xie L, Han J, Guo X. The application of deep learning in cancer prognosis prediction. *Cancers (Basel).* (2020) 12:603. doi: 10.3390/cancers12030603
33. Boeri C, Chiappa C, Galli F, De Berardinis V, Bardelli L, Carcano G, et al. Machine Learning techniques in breast cancer prognosis prediction: A primary evaluation. *Cancer Med.* (2020) 9:3234–43. doi: 10.1002/cam4.2811
34. Gao M, Guo Y, Xiao Y, Shang X. Comprehensive analyses of correlation and survival reveal informative lncRNA prognostic signatures in colon cancer. *World J Surg Oncol.* (2021) 19:104. doi: 10.1186/s12957-021-02196-4
35. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* (2020) 17:261–72. doi: 10.1038/s41592-020-0772-5
36. Reback J, McKinney W, jbrockmndel, Van den Bossche J, Augspurger T, Cloud P, et al. pandas-dev/pandas: Pandas 1.0.3 (Version v1.0.3). *Zenodo.* (2020). doi: 10.5281/zenodo.3715232
37. Davidson-Pilon. lifelines: survival analysis in Python. *J Open Source Software.* (2019) 4:1317. doi: 10.21105/joss.01317
38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* (2011) 12:2825–30. doi: 10.48550/arXiv.1201.0490
39. Chen T, Guestrin C. (2016). XGBoost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Vol. 16. pp. 785–94. doi: 10.1145/2939672.2939785
40. Shalev-Shwartz S, Ben-David S. *Understanding machine learning: from theory to algorithms*. USA: Cambridge University Press (2014). doi: 10.1017/CBO9781107298019
41. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. *Commun ACM.* (2016) 64:107–15. doi: 10.1145/3446776
42. Kawaguchi K, Bengio Y, Kaelbling L. Generalization in deep learning. In: Grohs P, Kutyniok G, editors. *Mathematical aspects of deep learning*. Cambridge University Press (2022). p. 112–48. doi: 10.1017/9781009025096.003
43. Sutskever I, Martens J, Dahl G, Hinton G. (2013). On the importance of initialization and momentum in deep learning, in: *Proceedings of the 30th International Conference on Machine Learning*, Vol. 28. pp. 1139–47. Available at: <https://proceedings.mlr.press/v28/sutskever13.html>.
44. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature.* (2011) 474:609–15. doi: 10.1038/nature10166
45. Gueli Alletti S, Capozzi VA, Rosati A, De Blasis I, Cianci S, Vizzielli G, et al. Laparoscopy vs. laparotomy for advanced ovarian cancer: a systematic review of the literature. *Minerva Med.* (2019) 110:341–57. doi: 10.23736/S0026-4806.19.06132-9
46. Tortorella L, Vizzielli G, Fusco D, Cho WC, Bernabei R, Scambia G, et al. Ovarian cancer management in the oldest old: improving outcomes and tailoring treatments. *Aging Dis.* (2017) 8:677–84. doi: 10.14336/AD.2017.0607
47. Becker T, Weberpals J, Jegg AM, So WV, Fischer A, Weisser M, et al. An enhanced prognostic score for overall survival of patients with cancer derived from a large real-world cohort. *Ann Oncol.* (2020) 31:1561–8. doi: 10.1016/j.annonc.2020.07.013
48. Scoles DR, Xu X, Wang H, Tran H, Taylor-Harding B, Li A, et al. Liver X receptor agonist inhibits proliferation of ovarian carcinoma cells stimulated by oxidized low density lipoprotein. *Gynecol Oncol.* (2010) 116:109–16. doi: 10.1016/j.ygyno.2009.09.034
49. Hou YF, Gao SH, Wang P, Zhang HM, Liu LZ, Ye MX, et al. $1\alpha,25(\text{OH})_2\text{D}_3$ suppresses the migration of ovarian cancer SKOV-3 cells through the inhibition of epithelial-mesenchymal transition. *Int J Mol Sci.* (2016) 17:1285. doi: 10.3390/ijms17081285
50. Lungchukiet P, Sun Y, Kasiappan R, Quarni W, Nicosia SV, Zhang X, et al. Suppression of epithelial ovarian cancer invasion into the omentum by $1\alpha,25$ -dihydroxyvitamin D3 and its receptor. *J Steroid Biochem Mol Biol.* (2015) 148:138–47. doi: 10.1016/j.jsbmb.2014.11.005
51. Ji M, Liu L, Hou Y, Li B. $1\alpha,25$ -Dihydroxyvitamin D3 restrains stem cell-like properties of ovarian cancer cells by enhancing vitamin D receptor and suppressing CD44. *Oncol Rep.* (2019) 41:3393–403. doi: 10.3892/or.2019.7116
52. Jin Q, Noel O, Nguyen M, Sam L, Gerhard GS. Bile acids upregulate BRCA1 and downregulate estrogen receptor 1 gene expression in ovarian cancer cells. *Eur J Cancer Prev.* (2018) 27:553–6. doi: 10.1097/CEJ.0000000000000398
53. Santoribio JD, Garcia-de la Torre A, Cañavate-Solano C, Arce-Matute F, Sanchez-del Pino MJ, Perez-Ramos S. Cancer antigens 19.9 and 125 as tumor markers in patients with mucinous ovarian tumors. *Eur J Gynaecol Oncol.* (2016) 37:26–9. Available at: <https://pubmed.ncbi.nlm.nih.gov/27048105/>.
54. Tanaka YO, Okada S, Satoh T, Matsumoto K, Oki A, Saida T, et al. Differentiation of epithelial ovarian cancer subtypes by use of imaging and clinical data: a detailed analysis. *Cancer Imaging.* (2016) 16:3. doi: 10.1186/s40644-016-0061-9
55. Zhang S, Cheng C, Lin Z, Xiao L, Su X, Zheng L, et al. The global burden and associated factors of ovarian cancer in 1990–2019: findings from the Global Burden of Disease Study 2019. *BMC Public Health.* (2022) 22:1455. doi: 10.1186/s12889-022-13861-y
56. Bakhru A, Buckanovich RJ, Griggs JJ. The impact of diabetes on survival in women with ovarian cancer. *Gynecol Oncol.* (2011) 121:106–11. doi: 10.1016/j.ygyno.2010.12.329
57. Kim SJ, Rosen B, Fan I, Ivanova A, McLaughlin JR, Risch H, et al. Epidemiologic factors that predict long-term survival following a diagnosis of epithelial ovarian cancer. *Br J Cancer.* (2017) 116:964–71. doi: 10.1038/bjc.2017.35
58. Becker S, Dossus L, Kaaks R. Obesity related hyperinsulinaemia and hyperglycaemia and cancer development. *Arch Physiol Biochem.* (2009) 115:86–96. doi: 10.1080/13813450902878054
59. Lugano R, Ramachandran M, Dimberg A. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cell Mol Life Sci.* (2020) 77:1745–70. doi: 10.1007/s00018-019-03351-7
60. Pennington KP, Walsh T, Harrell MI, Lee MK, Pennil CC, Rendi MH, et al. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin Cancer Res.* (2014) 20:764–75. doi: 10.1158/1078-0432.CCR-13-2287
61. Chalasani P, Livingston R. Differential chemotherapeutic sensitivity for breast tumors with “BRCAness”: a review. *Oncologist.* (2013) 18:909–16. doi: 10.1634/theoncologist.2013-0039
62. Kaye SB. Progress in the treatment of ovarian cancer—lessons from homologous recombination deficiency—the first 10 years. *Ann Oncol.* (2016) 27 Suppl 1:i1–3. doi: 10.1093/annonc/mdw082
63. Vanderstichele A, Busschaert P, Olbrecht S, Lambrechts D, Vergote I. Genomic signatures as predictive biomarkers of homologous recombination deficiency in ovarian cancer. *Eur J Cancer.* (2017) 86:5–14. doi: 10.1016/j.ejca.2017.08.029