IPVS 2025

37TH ANNUAL CONFERENCE OF THE INTERNATIONAL PAPILLOMAVIRUS SOCIETY

Research toward the global elimination of HPV-related diseases and cancers

OCTOBER 23-26, 2025 | BANGKOK, THAILAND

DON'T MISS THESE IMPORTANT DATES

30/07 • EARLY REGISTRATION DEADLINE

25/08 • LATE-BREAKING ABSTRACT SUBMISSION DEADLINE

ENJOY AN EXCLUSIVE 10% DISCOUNT

JGO

USE PROMO CODE

JGO10

ENTER THE PROMO CODE DURING THE REGISTRATION PROCESS

ipvconference.org | #IPVS2025

IPVS International Papillomavirus Society

J Gynecol Oncol. 2025 Jul;36(4):e53
https://doi.org/10.3802/jgo.2025.36.e53
pISSN 2005-0380·eISSN 2005-0399

JOURNAL OF GYNECOLOGIC ONCOLOGY

Original Article

Check for updates

Early prediction and risk stratification of ovarian cancer based on clinical data using machine learning approaches

Ting Gui,[1] **Dongyan Cao**,[1] **Jiaxin Yang**,[1] **Zhenhao Wei**,[2] **Jiatong Xie**,[2] **Wei Wang**,[2] **Yang Xiang**,[1] **Peng Peng**[1]

[1]Department of Obstetrics and Gynecology, National Clinical Research Center for Obstetric and Gynecologic Diseases, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China
[2]Goodwill Hessian Health Technology Co. Ltd, Beijing, People's Republic of China

OPEN ACCESS

Received: May 25, 2024
Revised: Oct 21, 2024
Accepted: Nov 11, 2024
Published online: Dec 17, 2024

Correspondence to
Peng Peng
Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, No. 1, Shuaifuyuan Road, Dongcheng District, Beijing 100730, People's Republic of China.
Email: pp_pengpeng2023@163.com
pengpeng@pumch.cn

© 2025. Asian Society of Gynecologic Oncology, Korean Society of Gynecologic Oncology, and Japan Society of Gynecologic Oncology
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs
Ting Gui
https://orcid.org/0000-0002-2265-2216
Dongyan Cao
https://orcid.org/0009-0004-1479-809X
Jiaxin Yang
https://orcid.org/0009-0002-7577-5671
Zhenhao Wei
https://orcid.org/0009-0003-7359-564X

ABSTRACT

Objective: Our study was aimed to construct a predictive model to advance ovarian cancer diagnosis by machine learning.
Methods: A retrospective analysis of patients with pelvic/adnexal/ovarian mass was performed. Potential features related to ovarian cancer were obtained as many as possible. The optimal machine learning algorithm was selected among six candidates through 5-fold cross validation. Top 20 features having the most powerful predictive significance were ranked by Shapley Additive Interpretation (Shap) method. Clinical validation was further performed to confirm whether our model could advance diagnosis of ovarian cancer.
Results: A total of 9,799 patients were collected. The inclusion criteria included age >18 years old, the first diagnosis being pelvic/adnexal/ovarian mass of undetermined significance, and pathological report indispensable. Four hundred and thirty-eight dimensional features were obtained after filtration. LightGBM showed the best performance with accuracy 88%. Among the top 20 features, 55% belonged to laboratory test report, 35% came from imaging examination report, and 10% were attributed to basic demographics and main symptom. Age, CA125, and risk of ovarian malignancy algorithm were the top three. Our predictive model performed stably in testing and clinical validation datasets, and was found to advance the diagnosis of ovarian cancer about 17 days before clinical pathological examination.
Conclusion: LightGBM was the optimal algorithm for our predictive model with accuracy of 88%. Laboratory test and imaging examination played essential roles in diagnosing ovarian cancer. Our model could advance the diagnosis of ovarian cancer before clinical pathological examination.

Keywords: Machine Prediction Methods; Risk Factors; Ovarian Cancer; Machine Learning; Predictive Learning Models

Generated by XmlInkpress

Jiatong Xie
https://orcid.org/0009-0000-0390-445X
Wei Wang
https://orcid.org/0009-0007-4548-4300
Yang Xiang
https://orcid.org/0009-0008-1700-9139
Peng Peng
https://orcid.org/0009-0001-8344-0356

Funding

This work was supported by the National High Level Hospital Clinical Research Funding (No. 2022-PUMCH-B-083).

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Author Contributions

Conceptualization: G.T., P.P.; Data curation: G.T., C.D., Y.J., X.Y.; Formal analysis: G.T., W.Z., X.J., W.W.; Funding acquisition: P.P.; Investigation: G.T., W.Z., X.J., W.W.; Methodology: G.T., W.Z., X.J., W.W., P.P.; Project administration: P.P.; Resources: C.D., Y.J., X.Y.; Software: W.Z., X.J., W.W.; Supervision: P.P.; Validation: W.Z., X.J., W.W.; Writing - original draft: G.T.; Writing - review & editing: C.D., Y.J., X.Y., P.P.

Synopsis

1. A predictive model for early prediction of ovarian cancer by machine learning was explored.
2. A total of 9,799 patients were collected for model training, testing, and validation.
3. Laboratory test and imaging examination were important in prediction.
4. Our predictive model could advance ovarian cancer diagnosis about 17 days before clinical pathological diagnosis.

INTRODUCTION

Ovarian cancer is relatively rare, accounting for 3.4% of all malignancies among females worldwide in 2020 [1]. It is a leading cause of cancer-attributable deaths and is the most fatal gynecological cancer with mortality rate reaching 4.7% [1]. Most patients were diagnosed at advanced stage, resulting in a poor long-term survival. Early diagnosis of ovarian cancer is difficult, generally attributed to the lack of symptoms [2]. Therefore, it is of great clinical significance to establish novel methods for early prediction and diagnosis, so that interventions like surgery and chemotherapy could be performed at an early stage improving the clinical outcome.

Machine learning, standing out as one of the most prominent artificial intelligence fields, has plenty of applications in almost every aspect of human life. Machine learning could be defined as computer algorithms that learn from past data to predict future data. The emergence of machine learning and big data provides us potentialities in differentiating benign tumor and malignancy through multiple clinical features.

Application of artificial intelligence in diagnosing ovarian cancer is already being taken from theory to clinical application [3-6]. However, previous studies mainly focused on one specific aspect, such as gene signature [7], tumor biomarkers [8], imaging characteristics [9]. These aspects alone fail to adequately account for the heterogeneity of clinical outcomes. Stratifying ovarian cancer patients by comprehensively fusing multiple aspects of clinical features is of pressing necessity [10].

Therefore, the main objectives of our study were 1) To construct a predictive model aimed at advancing ovarian cancer diagnosis; 2) To find the most significant risk predictors for ovarian cancer. To realize this, we applied machine learning on a comprehensive dataset. Specifically, medical records of patients with pelvic/adnexal/ovarian mass admitted to Peking Union Medical College Hospital in the recent 11 years were collected and analyzed retrospectively. Features having predictive potentiality of ovarian cancer were obtained as many as possible for model training & testing. Final clinical validation was performed to assess whether our model could advance diagnosis of ovarian cancer before pathological examination. It is hoped that construction of such predictive model could aid gynecologists in predicting and diagnosing ovarian cancer as early as possible.

MATERIALS AND METHODS

Our study was approved by the Ethics Committee of Peking Union Medical College Hospital (PUMCH) (IRB Number: I-22PJ808), who deemed written informed consent not necessary

due to the retrospective nature of the study. Privacy and security of data was paid fully attention, with all data de-privatized, including name, address, ID number, etc.

1. Data collection

The dataset was collected from the Medical Records Information System of Peking Union Medical College Hospital (PUMCH). The inclusion criteria included: 1) age >18 years old, 2) clinical diagnosis at the first visit of PUMCH being pelvic/adnexal/ovarian mass of undetermined significance; 3) pathological report indispensable. The exclusion criteria were clinical diagnosis at the first visit of PUMCH containing "cancer, carcinoma, sarcoma, cyst, myoma, benign, or malignant". Patient information included basic demographics, chief complaint, current and previous medical history, laboratory test reports, and imaging examination results.

At the first stage of model training & testing, 8,652 patients of pelvic/adnexal/ovarian mass of undetermined significance were collected primarily, among whom 1197 patients missed the pathological reports. Finally, 7,455 patients were included from 2013 January 1st to 2022 September 30th, with the ratio of ovarian cancer/non-ovarian cancer 1/1.35 (3,172/4,283).

At the second stage of model clinical validation, 2,721 patients of pelvic/adnexal/ovarian mass of undetermined significance were collected primarily, among whom 377 patients missed the pathological reports. Finally, 2,344 patients were enrolled from 2022 October 1st to 2023 September 30th, with the ration of ovarian cancer/non-ovarian cancer 1/1.28 (1,028/1,316).

In sum, a total of 9,799 patients of pelvic/adnexal/ovarian mass of undetermined significance with complete pathological reports were collected in our study. The basic information of patients was presented in **Table 1**. No significant differences were observed between the training & testing dataset and the validation dataset.

2. Feature selection

At the very beginning, a total of 27,971 dimensional features were extracted from 7,455 patients. To incorporate as many features as possible, the missing rate threshold was set as 99%, which meant features observed in less than 1% patients would be filtrated. After initial filtration, 663 dimensional features were obtained. Then features lack of obvious referential meaning or almost irrelevant to ovarian cancer were further filtered artificially.

Table 1. Basic information of patients

Basic characteristics	Model training & testing	Model validation	p-value
Number of patients	7,455	2,344	-
Age	49.04±14.95	48.92±12.35	0.082
Gravity	2.49±1.39	2.21±1.32	0.384
Parity	1.31±0.83	1.25±0.68	0.413
Menopause	9.44%	8.98%	0.543
FSH (IU/L)	38.06±34.98	37.11±32.44	0.062
Hemoglobin (g/L)	117.04±15.85	116.77±15.76	0.064
CA125 (IU/L)	548.03±1,591.09	612.08±1,725.33	0.061
ROMA (pre-menopause)	50.97±35.98	49.88±37.25	0.125
ROMA (post-menopause)	61.66±31.71	64.78±35.11	0.885
Systolic pressure (mmHg)	122.44±18.56	125.32±19.21	0.355
Diastolic pressure (mmHg)	75.58±11.20	72.64±14.38	0.211
Heart rate (beat per min)	85.28±13.51	86.44±15.74	0.074
Temperature (°C)	37.09±0.67	37.06±0.88	0.189
The ratio of ovarian cancer/Non-ovarian cancer	1:1.35 (3,174/4,281)	1:1.28 (1,028/1,316)	0.271

FSH, follicle stimulating hormone; ROMA, risk of ovarian malignancy algorithm.

https://doi.org/10.3802/jgo.2025.36.e53

Manual verification was performed for the deleted features to ensure that no clinically significant features were removed. Lastly, normalization was performed. Finally, a total of 438 dimensional features were obtained.

3. Data preprocessing

First the data was divided as discrete or continuous.

For discrete data, feature coding was performed by One-Hot method, putting 0 or 1 as the feature value, representing absence or presence. Missing discrete features were defaulted as absence.

For continuous data, outlier detection was performed based on 3σ principle of normal distribution. μ ± 3σ was the detection boundary, and values out of range were considered as abnormal and were substituted by boundary value. For missing values, KNN (K-Nearest Neighbor) was used to fill the blank (K=10). Based on clustering thought, assuming that similar samples having similar coordinates in high dimensional space, KNN method searches for K most similar samples through Euclidean Distance, and fills the blank with the average value of the similar samples.

To reduce the adverse effect of different dimensions and variances, the mean-standard deviation method was used for normalization by the following formula $z_{nor} = \frac{x - \mu}{s}$ (μ, average; s, standard deviation).

4. Machine learning

The dataset for model training & testing was randomly shuffled, 80% of the 7,455 cased selected for model training and the left 20% used for testing. Six machine learning algorithms were considered for model training, including SVM (Support Vector Machine), Bayes, LR (Logistic Regression), DT (Decision Tree), Light GBM (Light Gradient Boosting Machine), and XGBoost (eXtreme Gradient Boosting). The performance of six algorithms was compared through 5-fold cross validation, so as to select the most accurate algorithm for our predictive model.

5. Model performance evaluation

We used parametric grid search to optimize the parameters of each algorithm. Part of the parameters are as follows. In LR, tol=0.0001, C=0.5. In SVM, degree=5, tol=0.001, C=0.65. In Bayes, alph= 0.001, DT max_depth=39, min_samples_leaf=5. In XGBoost, learning_rate=0.2, n_estimators=112, gamma=1. In LightGBM, learning_rate=0.1, min_child_samples=20, n_estimators=100, num_leaves=42.

Evaluation parameters adopted in our study included accuracy, recall, precision, F1-Score, and receiver operating characteristic-area under the curve (AUC), based on the true positive (TP), false positive (FP), true negative (TN), and true false negative (FN). The relative formulae were as follows: Accuracy=$\frac{TP + TN}{TP + TN + FP + FN}$; Recall=$\frac{TP}{TP + FN}$; Precision=$\frac{TP}{TP + FP}$; F1_Score=$\frac{2 \cdot Precision \cdot ll}{Precision + Recall}$.

6. Feature importance evaluation

Machine learning model prediction is a complicated process. Shapley Additive Interpretation (Shap) method was used to explain the correlation between features and prediction results. The absolute value of Shap reflects the degree of influence on the model prediction results. The greater the absolute value, the greater the influence of the feature on the prediction. The positive or negative value of Shap reflects the direction of influence on the result. Shap

https://doi.org/10.3802/jgo.2025.36.e53

value >0 represents that features are tended to prompt the model to predict ovarian cancer; shap value <0 indicates the opposite result. Specifically in individual patient, after combining the supportive features and the non-supportive features, a final Shap value is obtained. If the value is >0, the Shap method tends to predict the patient having ovarian cancer.

7. Model clinical validation

After model training & testing, we further performed model clinical validation. The inclusion and exclusion criteria were stated above. Model prediction started from the first time of mass detection, and was performed at each subsequent clinical visit. Specifically, data of current visit and all the visits in the past 90 days were obtained. The data were converted into a vector containing 438 dimensional features, and the features were input into the trained model for prediction. The model output a probability value indicating whether the patient had ovarian cancer. The time interval between the clinical diagnosis confirmed by pathological report and the consistent prediction result provided by our model was calculated and analyzed. The performance of our predictive model in testing stage and clinical validation stage was compared to assess the model stability.

RESULTS

1. Model performance in training & testing

The performance of six algorithms in model training and testing was illustrated in **Table 2**. Parameters of Bayes had the lowest values. The results of Decision Tree were a little bit better. LR and SVM were much better but not the best. XGBoost and LightGBM were similar and better than the others.

Fig. 1 showed the ROC curve and its AUC value of six algorithms in testing dataset. The AUC of Bayes was a little bit higher than that of Decision Tree. The AUC of LR and SVM were similar, both higher than 0.93. XGBoost and LightGBM were the best, with AUC 0.9627 vs. 0.9675.

Therefore, LightGBM outperformed the other five algorithms, and was considered to be the optimal algorithm in predicting ovarian cancer in our study.

Table 2. Performance of 6 algorithms in model training and testing

Variables	Accuracy	Recall	Precision	F1-Score
Model training*				
LR	0.8601±0.0150	0.8260±0.0193	0.8325±0.0125	0.8277±0.0119
SVM	0.8418±0.0138	0.7919±0.0210	0.8485±0.0163	0.8247±0.0112
Bayes	0.7786±0.0155	0.7660±0.0381	0.8325±0.0163	0.7977±0.0095
Decision Tree	0.8110±0.0097	0.7734±0.0086	0.7714±0.0118	0.7724±0.0097
XGBoost	0.8683±0.0093	0.8554±0.0169	0.8786±0.0072	0.8667±0.0074
LightGBM	0.8883±0.0132	0.8765±0.0077	0.8569±0.0208	0.8665±0.0112
Model testing				
LR	0.8545	0.8180	0.8357	0.8467
SVM	0.8522	0.7944	0.8464	0.8404
Bayes	0.7989	0.8158	0.7961	0.8088
Decision Tree	0.8125	0.8134	0.8011	0.8142
XGBoost	0.8779	0.8670	0.8793	0.8731
LightGBM	0.8829	0.8766	0.8765	0.8786

DT, Decision Tree; Light GBM, Light Gradient Boosting Machine; LR, Logistic Regression; SVM, Support Vector Machine; XGBoost, eXtreme Gradient Boosting.
*Data are presented as mean ± standard deviation.

https://doi.org/10.3802/jgo.2025.36.e53

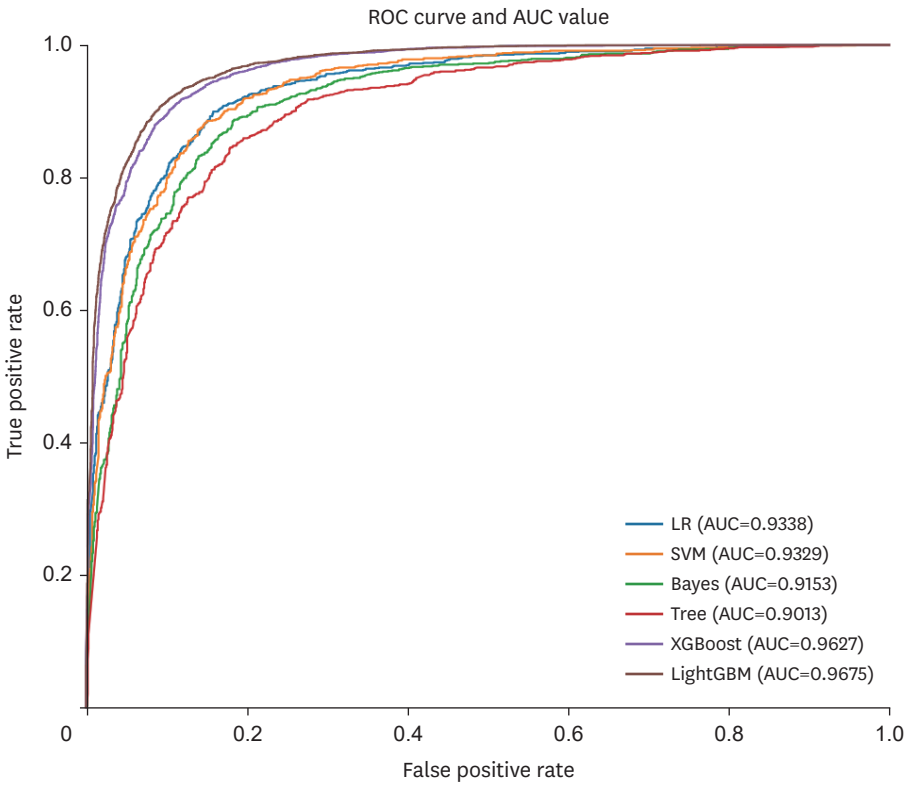

Fig. 1. The ROC curves and AUC values of six model learning algorithms in testing dataset.
AUC, area under the curve; Light GBM, Light Gradient Boosting Machine; LR, Logistic Regression; ROC, receiver operating characteristic; SVM, Support Vector Machine; XGBoost, eXtreme Gradient Boosting

2. Feature importance

Shap method was used to explain our predictive model. **Fig. 2A** illustrated the top 20 features having the most powerful predictive significance of ovarian cancer. The descending order indicated the degree of influence on the prediction result. Among the top 20 features, 55% belonged to laboratory test report, 35% came from imaging examination report, and 10% were attributed to basic demographics and main symptom. Age, CA125 and risk of ovarian malignancy algorithm (ROMA) ranked the top three.

Fig. 2B showed the direction of influence of top 20 features on the prediction of ovarian cancer. The older the age, the higher the tumor marker level, the lower the hemoglobin and albumin level, the higher the risk of developing ovarian cancer. If ultrasound found solid-cystic nodules or ascites, or magnetic resonance imaging showed abnormal signal/lesion or contrast enhancement, or CT indicated solid-cystic texture or contrast enhancement, the pelvic mass was prone to be ovarian cancer. The explanation of Shap was consistent with the clinical significance.

Fig. 2C represented the explaination of prediction result by Shap method in an individual patient. Red features represented forces supporting the prediction of ovarian cancer, while blue features, conversely, represented forces not supporting the prediction of ovarian cancer. After combining the forces of both sides, the final force value 0.73 was obtained. This patient was eventually predicted to have ovarian cancer.

https://doi.org/10.3802/jgo.2025.36.e53

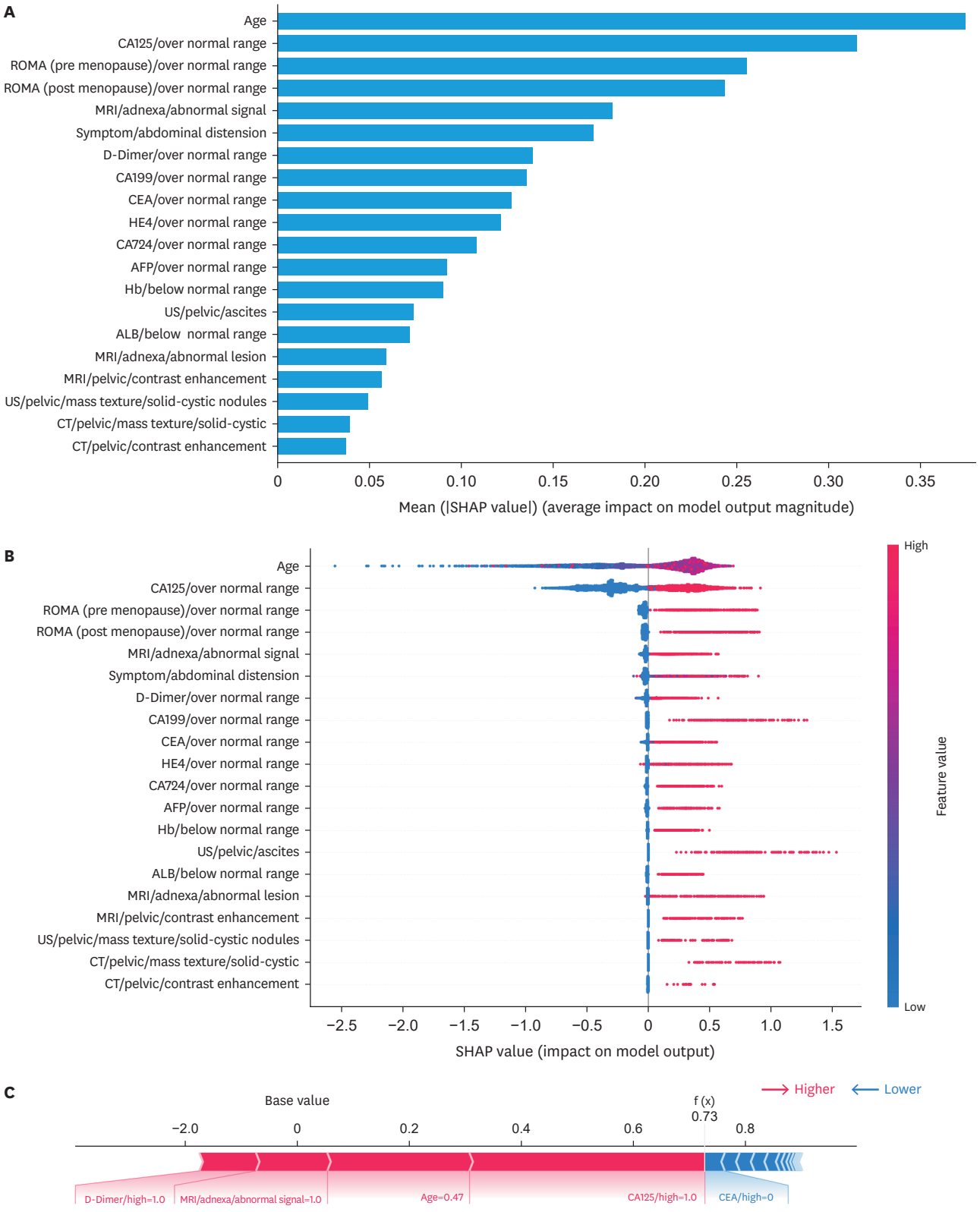

Fig. 2. Top 20 features having the most powerful predictive significance sorted by Shap method. (A) The degree of influence of Top 20 features on the prediction of ovarian cancer. (B) The direction of influence of Top 20 features on the prediction of ovarian cancer. (C) Interpretation of prediction result in individual patient by Shap method.
AFP, alpha fetoprotein; ALB, albumin; CEA, carcinoembryonic antigen; CT, computed tomography; Hb, hemoglobin; MRI, magnetic resonance imaging; ROMA, risk of ovarian malignancy algorithm; US, ultrasound.

https://doi.org/10.3802/jgo.2025.36.e53

3. Model clinical validation

Comparing the model performance in testing and clinical validation data set, our predictive model was found to be quite stable: accuracy 0.8829 vs. 0.8754, recall 0.8766 vs. 0.8853, precision 0.8765 vs. 0.8920, F1 score 0.8786 vs. 0.8854, and AUC 0.9675 vs. 0.9481.

By calculating the interval between the date of clinical pathological diagnosis and the date of our model prediction, we found that our model could advance the diagnosis of ovarian cancer about 16.98±22.73 days (ranging from 0 to 89 days) before pathological examination.

DISCUSSION

A total of 7,455 patients for model training & testing as well as 2,344 patients for model clinical validation were collected. 438 dimensional features covering multiple aspects were obtained after filtration. LightGBM was found to be the optimal for our predictive model after comparing the performance of six algorithm candidates. Top 20 features having the most powerful predictive significance were listed. Age, CA125 and ROMA ranked the top three. Our predictive model was found to be able to advance the diagnosis of ovarian cancer about 17 days before pathological examination.

Most studies concerning ovarian cancer prediction mainly focus on one specific feature, such as the tumor biomarker, ultrasound image, or gene signature profiling [7-9]. These results are relatively biased, since prediction or differentiation diagnosis of ovarian cancer is a comprehensive and complex process. Many predictive models have been developed to help with risk stratification, but none has been widely accepted or used in routine clinical practice. Lizuka et al. [11] compared the ROMA and Copenhagen Index (CPH-1) in their ability to distinguish ovarian cancer from benign ovarian tumors in Japanese women, finding ROMA and CPH-1 performed comparable well and better than CA125. However, ROMA and CPH-1 should be used with caution in practical situations, because their sensitivity were only only 54% and 55% [11]. The ultrasound based algorithms, such as the International Ovarian Tumor Analysis (IOTA), evaluating the risk of malignancy based on the ultrasound features, have been reported to have a good diagnostic accuracy [12]. However, when results are inconclusive, subjective assessment of ultrasonic findings by experienced sonographers should be performed additionally [12]. Multiple researches on early diagnosis of ovarian cancer based on mRNA, ctDNA, methylation, miRNA, and long non-coding RNA profiles have been reported [13-17]. Despite the alluring potential of high-through approaches, development and application of predictive models based on genomic or epigenomic profiling still has much room for improvement [18]. Risk stratification of ovarian cancer comprehensively considering multiple aspects of features remains poorly understood. Therefore, our study was aimed to construct a predictive model for early diagnosis based on comprehensive features covering demographics, chief complains, laboratory tests, and imaging examinations.

From the prospective of big data, large datasets are essential to improve the effect of machine learning. However, the lack of usable large datasets is one of the main challenges. In previous similar researches on ovarian cancer early diagnosis through artificial intelligence based on multiple features, the number of patients and features included was relatively small. Japanese researchers Akazawa et al. [4] performed a machine learning analysis to predict pathological diagnosis of ovarian cancer based on 202 patients with only 16 features. Lu et al. [19]

https://doi.org/10.3802/jgo.2025.36.e53

investigated the prediction performance of a machine learning model in a data set consisting of 349 patients with 49 features to classify benign ovarian tumor and ovarian cancer. In our study, a total of 9,799 patients of pelvic/adnexal/ovarian mass were collected and 438 features were obtained. As far as we know, our sample size and the number of features were the largest compared with previous similar researches.

Regarding the machine learning algorithm, LightGBM showed the best performance among six competing candidates with accuracy 88%. Our result was consistent with and better than previous researches. Akazawa et al. [4] used five machine learning algorithms, finding that the highest accuracy was 80% in XGBoost. Ahamad et al. [20] compared seven machine learning algorithms, and XGBoost achieved an accuracy of 86%. In addition to accuracy, LightGBM also presented very good performance in precision, recall, F1-score, and AUC score in our study. That is to say, our prediction result based on LightGBM was reasonable and reliable.

The prediction process of machine learning model is very complicated. Therefore, Shap method was used to explain the model based on Shapley values in game theory [21]. For individual patient, based on the Shap value of supportive and non-supportive features input into the model, the prediction result could be interpreted in a more visual way, making machine learning predictive models more transparent in clinical practice.

Among the top 20 features presented by Shap method, about half belonged to laboratory test report, and about one third belonged to imaging examination report. The significance of laboratory test features was higher than that of imaging features. This was probably because feature extraction was much more difficult for imaging examination reports. Compared with laboratory test reports, imaging examination reports consisted of descriptive words rather than concrete numbers. Subtle difference in descriptive words could affect the feature extraction.

Age, CA125 and ROMA ranked the top three risk predictors. Age is generally considered to be an important clinical indicator of ovarian cancer prognosis [10]. Currently, CA125 is the most widely used biomarker for ovarian cancer detection, but its use is limited by a low specificity. Multi-marker panels have been developed by combining molecular biomarkers to improve the diagnostic efficacy. Based on extensive scientific evidence, CA125 combined with HE4 to form the ROMA, has become widespread in clinical practice in evaluation of adnexal masses with improved sensitivity and specificity [22,23]. Our results were consistent with clinical practice and previous researches.

Machine learning is capable of finding hidden patterns in data by examining a collection of characteristics. Some features, imperceptible to traditional methods but having strong clinical association, could probably be found by machine learning. In our study, CA724 was found to be ranked the 11th in top 20. Although detection of CA724 is not as routine as CA125 in clinical practice, many studies have shown that CA724 is dramatically elevated in ovarian cancer and is a specific tumor marker for risk assessment of unique clear-cell carcinoma subtype [24-26].

The purpose of constructing such a predictive model was to apply it in clinical practice of ovarian cancer diagnosis. By comparing the performance of our model in testing vs. clinical validation dataset, our model was found to be quite stable. The values of five parameters, accuracy, recall, precision, F1 score and AUC, were all very close. Furthermore, it is worth

mentioning that, our model could advance the ovarian cancer diagnosis nearly 17 days. This meant that using our predictive model could probably enable patients to receive clinical treatment much earlier than before.

One strength of our study was the single institution uniform management, which enabled us to implement model clinical validation based on preliminary training & testing. Another strength was the relatively large size of dataset and features compared with other researches.

The major weaknesses included: 1) Prospective studies had not yet performed to minimize bias and to validate findings. 2) Generalizability needed to be enhanced by multi-center and multi-regional data. 3) External validation was required to confirm the model's applicability across different settings. 4) Model interpretability should be improved to facilitate clinical adoption.

In addition, although we set appropriate inclusion conditions, careful data preprocessing, and additional manual verification, there might still be some noise and inaccuracies in the data. However, through comprehensive consideration of a variety of evaluation metrics, the machine learning model developed in this study achieved good results in both testing dataset and validation dataset, presenting good robustness and generalization ability. This indicates that artificial intelligence model has strong adaptability and potential to deal with the complex and changeable data environment in the real world.

Classification of benign and malignant pelvic mass has profound effect on the prognosis of patients. It is of paramount clinical significance to improve the accuracy of early diagnosis of ovarian cancer. Machine learning in ovarian cancer prediction is a growing field with great potential. Although the prediction accuracy could not reach 100%, machine learning prediction could aid the gynecologists to diagnose and make treatment decisions. Next step is to perform prospective multi-center studies, so as to minimize bias and enhance generalizability.

LightGBM was the optimal algorithm for our predictive model with an accuracy of 88%. Laboratory test and imaging examination played essential roles in diagnosing ovarian cancer. Age, CA125, and ROMA were the top three risk predictors strongly related to ovarian cancer. Our model could advance the diagnosis of ovarian cancer before clinical pathological examination.

ACKNOWLEDGEMENTS

Thank the entire medical staff of the gynecological cancer center of Peking Union Medical College Hospital for their devotion to data collection. Thank engineer Bingqing Yang from Goodwill Hessian Health Technology Co. Ltd for her work in verification of data, feature importance, and modeling.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209-49. PUBMED | CROSSREF

2. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin 2018;68:284-96. PUBMED | CROSSREF
3. Kawakami E, Tabata J, Yanaihara N, Ishikawa T, Koseki K, Iida Y, et al. Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. Clin Cancer Res 2019;25:3006-15. PUBMED | CROSSREF
4. Akazawa M, Hashimoto K. Artificial intelligence in ovarian cancer diagnosis. Anticancer Res 2020;40:4795-800. PUBMED | CROSSREF
5. Zhou J, Cao W, Wang L, Pan Z, Fu Y. Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer. Comput Biol Med 2022;146:105608. PUBMED | CROSSREF
6. Wang Y, Lin W, Zhuang X, Wang X, He Y, Li L, et al. Advances in artificial intelligence for the diagnosis and treatment of ovarian cancer (review). Oncol Rep 2024;51:46. PUBMED | CROSSREF
7. Bahado-Singh RO, Ibrahim A, Al-Wahab Z, Aydas B, Radhakrishna U, Yilmaz A, et al. Precision gynecologic oncology: circulating cell free DNA epigenomic analysis, artificial intelligence and the accurate detection of ovarian cancer. Sci Rep 2022;12:18625. PUBMED | CROSSREF
8. Mikdadi D, O'Connell KA, Meacham PJ, Dugan MA, Ojiere MO, Carlson TB, et al. Applications of artificial intelligence (AI) in ovarian cancer, pancreatic cancer, and image biomarker discovery. Cancer Biomark 2022;33:173-84. PUBMED | CROSSREF
9. Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou Y, et al. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. EClinicalMedicine 2022;53:101662. PUBMED | CROSSREF
10. Boehm KM, Aherne EA, Ellenson L, Nikolovski I, Alghamdi M, Vázquez-García I, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. Nat Can 2022;3:723-33. PUBMED | CROSSREF
11. Iizuka M, Hamada Y, Matsushima J, Ichikawa T, Irie T, Yamaguchi N, et al. Comparison of the risk of ovarian malignancy algorithm and Copenhagen Index for the preoperative assessment of Japanese women with ovarian tumors. J Obstet Gynaecol Res 2023;49:2717-27. PUBMED | CROSSREF
12. Dakhly DM, Gaafar HM, Sediek MM, Ibrahim MF, Momtaz M. Diagnostic value of the International Ovarian Tumor Analysis (IOTA) simple rules versus pattern recognition to differentiate between malignant and benign ovarian masses. Int J Gynaecol Obstet 2019;147:344-9. PUBMED | CROSSREF
13. Terp SK, Stoico MP, Dybkær K, Pedersen IS. Early diagnosis of ovarian cancer based on methylation profiles in peripheral blood cell-free DNA: a systematic review. Clin Epigenetics 2023;15:24. PUBMED | CROSSREF
14. Davies M, Davey MG, Miller N. The potential of microRNAs as clinical biomarkers to aid ovarian cancer diagnosis and treatment. Genes (Basel) 2022;13:2054. PUBMED | CROSSREF
15. Oncul S, Amero P, Rodriguez-Aguayo C, Calin GA, Sood AK, Lopez-Berestein G. Long non-coding RNAs in ovarian cancer: expression profile and functional spectrum. RNA Biol 2020;17:1523-34. PUBMED | CROSSREF
16. Xia T, Fang C, Chen Y. Advances in application of circulating tumor DNA in ovarian cancer. Funct Integr Genomics 2023;23:250. PUBMED | CROSSREF
17. Cammarata G, Barraco N, Giusti I, Gristina V, Dolo V, Taverna S. Extracellular vesicles-ceRNAs as ovarian cancer biomarkers: looking into circRNA-miRNA-mRNA code. Cancers (Basel) 2022;14:3404. PUBMED | CROSSREF
18. Govindarajan M, Wohlmuth C, Waas M, Bernardini MQ, Kislinger T. High-throughput approaches for precision medicine in high-grade serous ovarian cancer. J Hematol Oncol 2020;13:134. PUBMED | CROSSREF
19. Lu M, Fan Z, Xu B, Chen L, Zheng X, Li J, et al. Using machine learning to predict ovarian cancer. Int J Med Inform 2020;141:104195. PUBMED | CROSSREF
20. Ahamad MM, Aktar S, Uddin MJ, Rahman T, Alyami SA, Al-Ashhab S, et al. Early-stage detection of ovarian cancer based on clinical data using machine learning approaches. J Pers Med 2022;12:1211. PUBMED | CROSSREF
21. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 31st International Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec 4-9; Long Beach, CA, USA. Red Hook, NY: Curran Associates Inc.; 2017.
22. Zhang R, Siu MK, Ngan HY, Chan KK. Molecular biomarkers for the early detection of ovarian cancer. Int J Mol Sci 2022;23:12041. PUBMED | CROSSREF
23. Dochez V, Caillon H, Vaucel E, Dimet J, Winer N, Ducarme G. Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. J Ovarian Res 2019;12:28. PUBMED | CROSSREF

24. Li M, Men X, Zhang X. Diagnostic value of carbohydrate antigen 72-4 combined with carbohydrate antigen 15.3 in ovarian cancer, cervical cancer and endometrial cancer. J BUON 2020;25:1918-27. PUBMED
25. Wang Q, Wu Y, Zhang H, Yang K, Tong Y, Chen L, et al. Clinical value of serum HE4, CA125, CA724, and ROMA index for diagnosis of ovarian cancer and prediction of postoperative recurrence. Clin Lab 2019;65: PUBMED | CROSSREF
26. Xin B, Ji KQ, Liu YS, Zhao XD. NFAT overexpression correlates with CA724 and poor prognosis of ovarian clear-cell carcinoma subtype. Reprod Sci 2021;28:745-56. PUBMED | CROSSREF