

Original Article



OPEN ACCESS

Received: Dec 11, 2018

Revised: Feb 7, 2019

Accepted: Feb 9, 2019

Correspondence to

Sung Wook Seo

Department of Orthopedic Surgery, Samsung Medical Center and Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea.
E-mail: sungwseo@skku.edu

*The first two authors contributed equally to this study.

Copyright © 2019. Asian Society of Gynecologic Oncology, Korean 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

E Sun Paik <https://orcid.org/0000-0002-6172-5860>
Jeong-Won Lee <https://orcid.org/0000-0002-6110-4909>
Jeong-Yeol Park <https://orcid.org/0000-0003-2475-7123>
Ju-Hyun Kim <https://orcid.org/0000-0003-4007-5739>
Mijung Kim <https://orcid.org/0000-0003-3995-5384>
Tae-Joong Kim <https://orcid.org/0000-0002-9693-9164>
Chel Hun Choi <https://orcid.org/0000-0002-0199-6669>

Prediction of survival outcomes in patients with epithelial ovarian cancer using machine learning methods

E Sun Paik ^{1,*}, **Jeong-Won Lee** ^{1,2,*}, **Jeong-Yeol Park** ³, **Ju-Hyun Kim** ³, **Mijung Kim** ⁴, **Tae-Joong Kim** ¹, **Chel Hun Choi** ¹, **Byoung-Gie Kim** ¹, **Duk-Soo Bae** ¹, **Sung Wook Seo** ^{2,5}

¹Department of Obstetrics and Gynecology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

²Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University School of Medicine, Seoul, Korea

³Department of Obstetrics and Gynecology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

⁴Department of Electronics and Information System, Ghent University, Ghent, Belgium

⁵Department of Orthopedic Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

ABSTRACT

Objectives: The aim of this study was to develop a new prognostic classification for epithelial ovarian cancer (EOC) patients using gradient boosting (GB) and to compare the accuracy of the prognostic model with the conventional statistical method.


Methods: Information of EOC patients from Samsung Medical Center (training cohort, n=1,128) was analyzed to optimize the prognostic model using GB. The performance of the final model was externally validated with patient information from Asan Medical Center (validation cohort, n=229). The area under the curve (AUC) by the GB model was compared to that of the conventional Cox proportional hazard regression analysis (CoxPHR) model.

Results: In the training cohort, the AUC of the GB model for predicting second year overall survival (OS), with the highest target value, was 0.830 (95% confidence interval [CI]=0.802–0.853). In the validation cohort, the GB model also showed high AUC of 0.843 (95% CI=0.833–0.853). In comparison, the conventional CoxPHR method showed lower AUC (0.668 (95% CI=0.617–0.719) for the training cohort and 0.597 (95% CI=0.474–0.719) for the validation cohort) compared to GB. New classification according to survival probability scores of the GB model identified four distinct prognostic subgroups that showed more discriminately classified prediction than the International Federation of Gynecology and Obstetrics staging system.

Conclusion: Our novel GB-guided classification accurately identified the prognostic subgroups of patients with EOC and showed higher accuracy than the conventional method. This approach would be useful for accurate estimation of individual outcomes of EOC patients.

Keywords: Machine Learning; CA-125 Antigen; Ovarian Neoplasms; Prognosis; Survival

Byoung-Gie Kim 
<https://orcid.org/0000-0002-0572-8450>

 Duk-Soo Bae 
<https://orcid.org/0000-0003-0016-1704>

 Sung Wook Seo 
<https://orcid.org/0000-0001-8643-7825>

Funding

This study was supported by grants from the National R&D Program for Cancer Control, Ministry for Health, Welfare and Family Affairs, Republic of Korea (1520100); the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education and Science Technology (MEST) (2016R1A2B3006644); the NRF grant funded by the Ministry of Science and ICT (MSIP) (2016R1A5A2945889); the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI14C3418); Basic Science Research Program through the NRF funded by the MEST (2016R1E1A1A0191433); Samsung Medical Center grant (OTC1190111).

Conflict of Interest

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

Author Contributions

Conceptualization: L.J.W., S.S.W.; Data curation: P.E.S., P.J.Y., K.J.H., K.M.; Formal analysis: S.S.W.; Investigation: P.E.S., L.J.W., S.S.W.; Methodology: P.E.S., K.M., S.S.W.; Software: K.M., S.S.W.; Supervision: L.J.W., S.S.W.; Validation: P.J.Y., K.J.H., K.M., S.S.W.; Visualization: P.E.S., S.S.W.; Writing - original draft: P.E.S.; Writing - review & editing: K.T.J., C.C.H., K.B.G., B.D.S., S.S.W.

INTRODUCTION

Epithelial ovarian cancer (EOC) is a leading cause of death among patients who suffer from gynecological malignancies [1]. EOC presents at a late stage in most cases, and 40%–60% of all patients and 75% of patients at an advanced stage will eventually experience recurrence [2,3]. The accurate estimation of survival for EOC patients is important because prognosis could be a key determinant of treatment aggressiveness.

Previously known cancer-related risk factors for EOC patients include International Federation of Gynecology and Obstetrics (FIGO) stage, histologic type, residual disease status after surgery, histologic grade, and cancer antigen 125 (CA-125) level [4-7]. FIGO stage, which had been recently revised in 2013, is one of the most frequently used indicators for the prognosis of EOC patients. In the revised FIGO stage, the number of substages increased for more specific prognostications. However, due to the small number of each group, the substaging decreased the statistical power, which drew some criticism because this system eventually increased its uncertainty [4,8].

The gradient boosting (GB) algorithm is one of the most powerful machine learning techniques that has exhibited success in clinical applications. It is one of the ensemble methods that rely on combining a large number of weak simple predictors to obtain a stronger prediction. Therefore, the GB model is highly flexible in analyzing diverse patterns of data. GB algorithms have been applied to solve many different medical problems such as prediction of acute lower gastrointestinal bleeding and epilepsy using electroencephalogram data [9,10]. By applying the GB method for survival prediction, time sequential information could be used for more accurate results. In EOC, the serum level of CA-125 fluctuates according to therapeutic response of patients in addition to individual variations [11,12]. Since conventional statistical models are limited to analyzing the time sequence patterns of CA-125 level with other individual variables, we could use the GB method to analyze the data to increase the accuracy of prediction.

In this study, we aimed to develop a new prognostic classification for predicting the survival of EOC patients using machine learning technology, GB. The current study also evaluated the accuracy of the prediction model with the GB method by comparing it with the conventional statistical method.

MATERIALS AND METHODS

1. Patients

For training cohort, data of all EOC patients who underwent primary treatment at Samsung Medical Center between January 2002 and December 2015 were included. We performed a retrospective review from the electric medical records. Data of 1,128 EOC patients were analyzed for the training cohort. For external validation, we used data from EOC patients who were treated in Asan Medical Center from 2004 to 2010 with sufficient clinical information as training cohort, and data of 229 patients were analyzed (**Fig. 1**). This study was approved by the Institutional Review Board (IRB No. of Samsung Medical Center, 2016-08-090 and Asan Medical Center, 2017-0500).

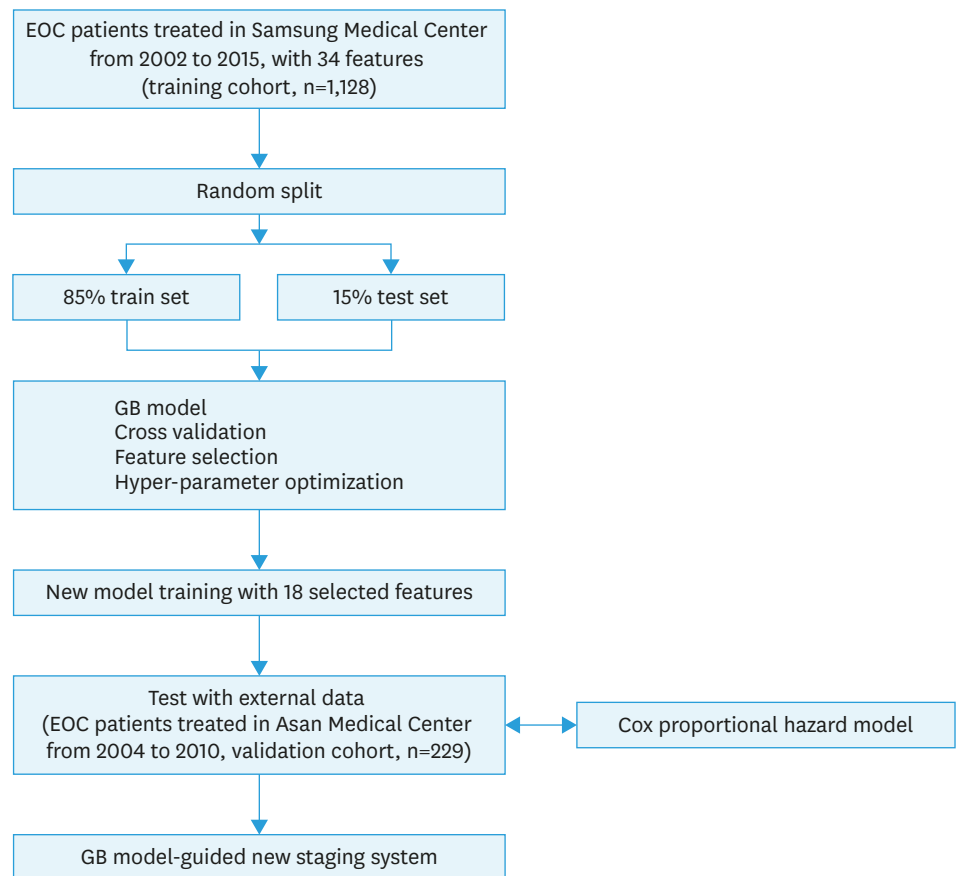


Fig. 1. Flowchart of development of the prediction model for overall survival in ovarian cancer patients. EOC, epithelial ovarian cancer; GB, gradient boosting.

2. Treatment and follow-up

Primary surgical treatment of EOC consisted of hysterectomy, bilateral salpingo-oophorectomy, omentectomy, retroperitoneal (pelvic and para-aortic) lymphadenectomy, and any tumorectomy of metastatic lesions if applicable. Peritoneal washing was routinely carried out. If any abnormalities were identified, peritoneal biopsies from different sites were appropriately performed. After debulking surgery, patients started the first cycle of platinum-based combination chemotherapy, which was repeated every three weeks for 6 cycles. CA-125 measurements were performed at each cycle of chemotherapy. Abdominopelvic computed tomography (CT) scan was routinely performed after first three cycles of chemotherapy and after 6 cycles of first line treatment. Neoadjuvant chemotherapy was considered for patients with bulky stage II I to IV disease who were unlikely to be completely cytoreduced to no residual disease, or patients who are poor surgical candidates.

After primary treatment, patients were assessed by physical examination, complete blood count, and chemistry with serum tumor markers, including CA-125 measurements, every three months for the first 2 years and twice per year thereafter. Chest radiography and abdominopelvic CT scan (or alternatively abdominopelvic magnetic resonance imaging) were performed every 6 months for the first 3 years and every 12 months thereafter. Additional diagnostic procedures were performed according to specific clinical suspicions. If recurrence was suspected with symptoms or CA-125 elevation, additional imaging studies were performed. Recurrence may have been

detected by imaging studies with or without CA-125 elevation. Response to chemotherapy was assessed and recorded according to Response Evaluation Criteria in Solid Tumors criteria.

3. Variables for analysis

Thirty-four variables of EOC patients were analyzed (**Supplementary Table 1**). Each variable was shown as continuous, or assigned to binary/multiple categories for descriptive purposes and statistical analysis. Age, FIGO stage, histology, grade, estimated blood loss (EBL) at primary surgery, American Society of Anesthesiologist (ASA) status for preoperative evaluation, operation site, lymphatic metastasis, and interval from surgery to chemotherapy (ISC) were obtained from the medical records. Additionally, we used the serial serum CA-125 level during primary treatment (pre-operative, post-operative, and first to sixth cycles). For describing characteristic of patients group, summary statistics were used to describe the data. Medians (range) or means (standard deviation) were used for continuous variables. After the Shapiro-Wilks test confirmed normal distributions, Mann-Whitney U test was used to compare median values and Student's t-test was used to compare mean values. Categorical variables were presented as frequencies (percentages). Fisher's exact test or χ^2 test was used to analyze the distribution of characteristics. All p-values were 2-sided, and we considered p-values less than 0.05 as statistically significant.

4. Strategy for data separation and feature selection for development of the GB model

A flowchart for the construction of a new prediction model for overall survival (OS) is shown in **Fig. 1**. Patients from Samsung Medical Center (n=1,128) were randomly sorted into a training set (85%) and a test set (15%). Missing data were imputed using the k-nearest neighbor algorithm [13] after separating the training and test sets. The feature values were preprocessed by normalization. Five-fold cross-validation was performed with the training data to optimize the hyper-parameters of the GB machine. The base learner of our model was decision tree where the maximum depth was 6, and Friedman's mean square error was used as a decision rule in each node. GB generates an additive learner in every forward stage. In each stage, regression trees are fit by gradient descent method. In this study, the model was boosted by 1,000 stages. Learning rate which controls the contribution of each tree was limited to 0.03.

The prediction accuracy of the GB model was evaluated over the test set at the discrete time point (from the first to the fifth years after the first operation) for selection of target year. After target year with highest accuracy was selected, the area under the curve (AUC) of the GB machine was evaluated as the number of variables was reduced in order to find a model trained with optimal number of variables. Models were generated using train set with different number of variables, and AUC of the trained model was evaluated using test set. The distribution of probable AUC values in each model was generated using bootstrap resampling, where training sets (85%) and their corresponding test sets (15%) were resampled 100 times. The median AUC of each model was compared with those of other models using Wilcoxon test, and we choose an optimal model whose AUC is significantly higher than those of the others. The model was retrained using the optimal number of variables and then evaluated with an external dataset from the other institute (validation cohort, n=229).

5. Performance evaluation of the GB model by comparing with conventional cox proportional hazards model, and Kaplan-Meier curve analysis for subgroups by GB model

A Cox proportional hazards regression analysis (CoxPHR) for prediction of target-year survival was generated using the same train set for finding significant covariates for the final

CoxPHR model development. We evaluated the accuracy of the CoxPHR model for predicting target-year prognosis after surgery using the train set. External validation of the CoxPHR model and GB model was performed over the validation dataset from the other hospital. Patients of external validation group were evenly divided to 4 subgroups (A, B, C, and D) according to the survival probability scores predicted by the GB model. Kaplan-Meier curve analysis was performed for 4 subgroups (A, B, C, and D), and then compared with Kaplan-Meier curves by FIGO stage (I, II, III, and IV) in test set. These analysis was executed using SAS version 9.4 (SAS Institute, Cary, NC, USA) and R 3.3.2 (Vienna, Austria; <http://www.R-project.org/>). AUCs by GB model and conventional cox proportional hazards model were compared using non-parametric statistical tests described by Hanley and McNeil using the MedCalc program (version 12.7; MedCalc Software, Ostend, Belgium) [14]. The Scikit-learn library was used to generate the GB model and pre-process other data [15].

RESULTS

1. Demographic analysis

Patient demographic and clinical characteristics are listed in **Table 1**. We analyzed the medical records of 1,128 patients with EOC treated at Samsung Medical Center as the training cohort and of 229 patients treated at Asan Medical Center as the validation cohort. There were 607 cases of cancer recurrence and 416 cases of overall death in the median follow-up of 47 months (range of 4 to 177 months) for the training cohort. The majority of patients had FIGO stage III (56.6%) and grade 3 disease (56.8%) of high-grade serous type (61.6%). No residual disease after initial debulking surgery was observed in 469 patients (41.5%). For the validation cohort, there were 143 cases of cancer recurrence and 81 cases of overall death in the median follow-up of 63 months (range of 6 to 143 months). There were no statistically significant differences in age, FIGO stage, or histology between the training and validation cohorts. However, differences in grade, residual disease status after debulking surgery, ASA score, lymphatic invasion status, type of treatment, ISC, and post-operative CA-125 level were observed between the 2 groups.

2. Target year selection and important feature selection for development of the GB model

The evaluation of prediction accuracy of GB model over the train set at the discrete time point revealed that the prediction accuracy was highest at the second year (**Supplementary Fig. 1**). The median AUC of survival prediction at the second year was 0.74 (95% confidence interval [CI]=0.71–0.79). Therefore, we used the second-year survival data as the target year upon which the GB model was trained and validated through 100 bootstrap resampling repetitions with a gradual reduction in the number of variables. A list of the variables used in the analysis with descriptions is provided in **Supplementary Table 1**, and variables (n=34) by relative rank or importance of variables (result of quantifying the association with second-year OS) are shown in **Fig. 2**. As a result of Wilcoxon test, the highest AUC is shown at the 18th variable (0.830; 95% CI=0.802–0.853) (**Fig. 3**). The 18 selected variables are shown in top in **Fig. 2**, and the AUC of the final model validated with resampling repetitions over those features is shown in **Supplementary Fig. 2**.

3. Cox proportional hazards model for predicting second year OS probability

All variables of training cohort were analyzed with CoxPHR for second year OS. The CoxPHR analysis found six significant covariates for second year OS (age, cell grade, histology,

Table 1. Patient demographics and clinical characteristics of the training cohort and test cohort

Characteristics	Training set (n=1,128)	Test set (n=229)	p-value
Age (yr)	52 (18–88)	50 (19–79)	0.172
Stage			0.406
I	265 (23.5)	53 (23.1)	
II	109 (9.7)	18 (7.9)	
III	638 (56.6)	141 (61.6)	
IV	116 (10.3)	17 (7.4)	
Histology			0.540
High-grade serous	695 (61.6)	143 (62.4)	
Low-grade serous	26 (2.3)	4 (1.7)	
Endometrioid	118 (10.4)	20 (8.7)	
Clear cell	107 (9.5)	18 (7.9)	
Mucinous	107 (9.5)	20 (8.7)	
Mixed	34 (3.0)	11 (4.8)	
Others	42 (3.7)	13 (5.7)	
Grade (%)			<0.001
1	105 (9.3)	20 (8.7)	
2	233 (20.6)	41 (17.9)	
3	641 (56.8)	105 (45.9)	
Unknown	149 (13.3)	63 (27.5)	
Residual disease after surgical debulking (cm)			<0.001
No gross residual	569 (50.4)	109 (47.6)	
≤1	280 (24.8)	82 (35.8)	
>1	280 (24.8)	38 (16.6)	
ASA score			<0.001
1	515 (45.6)	186 (81.2)	
2	424 (37.6)	42 (18.3)	
3	41 (3.6)	1 (0.04)	
Unknown	148 (13.2)	0	
Lymphatic metastasis status			
Pelvic LN metastasis	239 (21.2)	100 (43.7)	<0.001
Paraortic LN metastasis	147 (13.0)	42 (18.3)	<0.001
Lymph node dissection rate			
Pelvic LN	736 (65.2)	201 (87.8)	<0.001
Paraortic LN	388 (34.4)	187 (81.7)	<0.001
ISC (day)	9 (2–92)	16 (4–67)	<0.001
Type of primary treatment			<0.001
Primary debulking surgery+adjuvant chemotherapy	974 (86.3)	220 (96.1)	
Neoadjuvant chemotherapy+interval debulking surgery	81 (7.2)	0	
Primary debulking surgery	74 (6.6)	9 (3.9)	
Type of chemotherapy regimen			<0.001
Taxane-carboplatin	922 (81.7)	174 (76.0)	
Taxane-cisplatin	85 (7.5)	7 (3.1)	
Cyclophosphamide-cisplatin	27 (2.4)	4 (1.7)	
Other	95 (8.4)	44 (19.2)	
CA-125 (U/mL)			
Pre-operative	431.5 (0.1–10,080.0)	448.0 (3.6–31,600.0)	0.383
Post-operative	117.2 (1.1–11,390.0)	103.5 (3.8–5,080.0)	0.001
1st cycle of chemotherapy	30.8 (0.0–7,724.0)	43.3 (4.0–2,540.0)	0.307
2nd cycle of chemotherapy	13.8 (0.1–8,389.0)	19.5 (3.0–4,837.0)	0.782
3rd cycle of chemotherapy	10.3 (0.0–4,334.0)	12.6 (3.1–957.0)	0.535
4th cycle of chemotherapy	9.0 (0.1–5,035.0)	11.1 (2.5–388.0)	0.138
5th cycle of chemotherapy	7.9 (0.1–3,027.0)	9.3 (2.3–1,260.0)	0.373
6th cycle of chemotherapy	7.4 (0.1–4,419.2)	8.2 (2.4–1,850.0)	0.674

Values are presented as number of patients (%) or median (range).

ASA, American Society of Anesthesiologists; CA-125, cancer antigen 125; ISC, interval from surgery to chemotherapy; LN, lymph node.

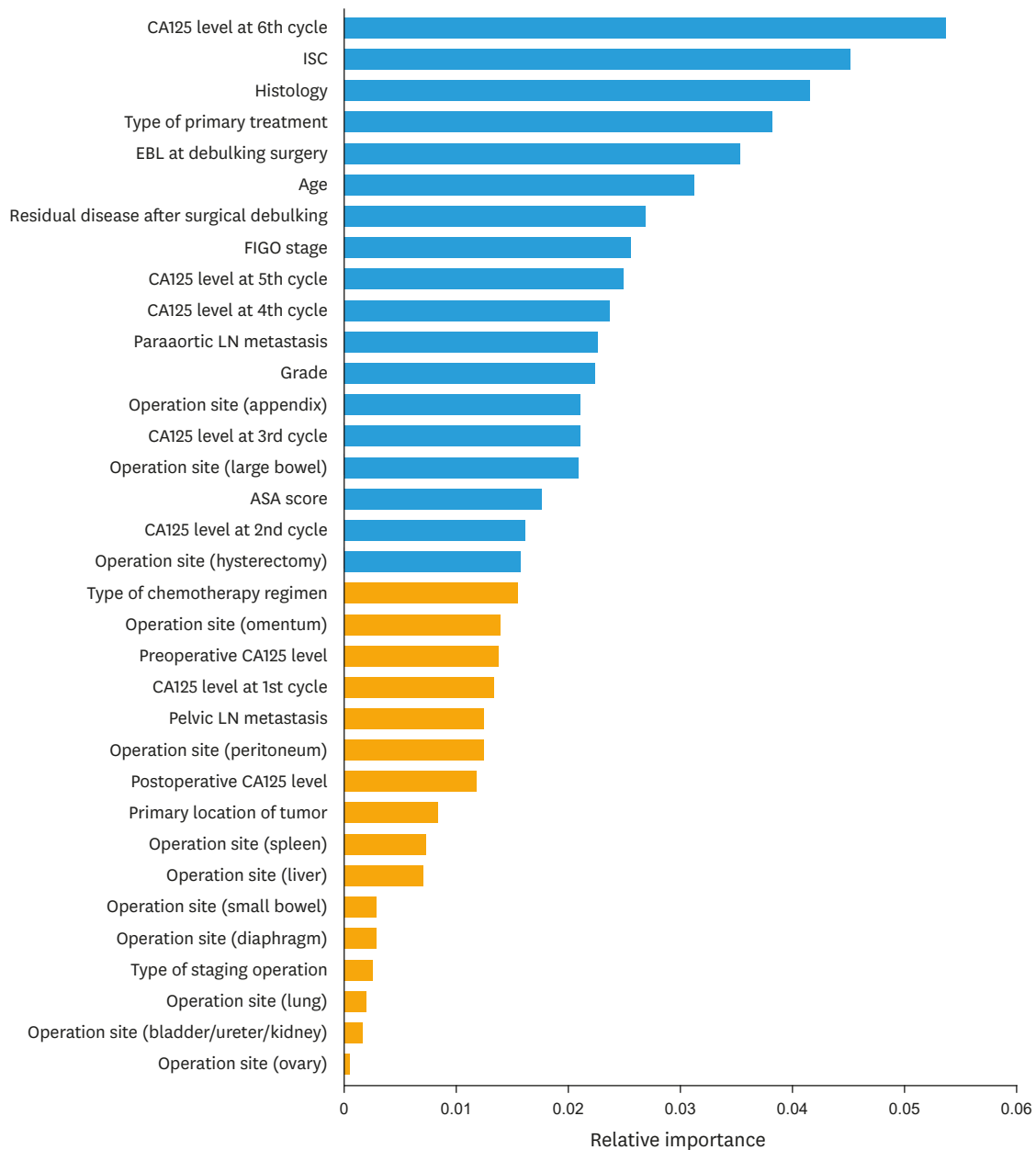


Fig. 2. Bar graph showing clinical features by relative rank of variables to second year overall survival. ASA, American Society of Anesthesiologist; CA-125, cancer antigen 125; EBL, estimated blood loss; FIGO, International Federation of Gynecology and Obstetrics; ISC, interval from surgery to chemotherapy; LN, lymph node.

postoperative CA-125 level, lymphatic metastasis in para-aortic lymph node, and estimated blood loss during surgery) (**Supplementary Table 2**). In the internal validation, the AUC of the receiver operating characteristic curves of the second year OS probability with CoxPHR model was 0.668 (95% CI=0.617–0.719) with six covariates (**Supplementary Fig. 3**).

4. External validation and performance comparison in predicting second year OS probability

The performances of the CoxPHR model and the GB model were finally tested with the external validation cohort. The AUC of the second year OS probability was 0.597 (95%

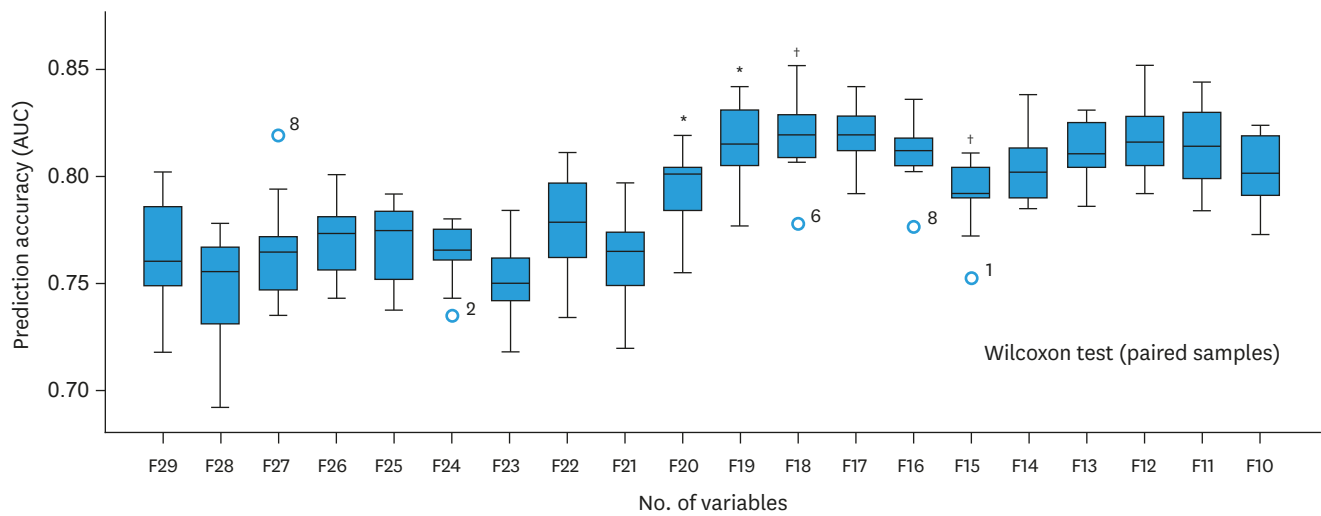


Fig. 3. Comparison of the median AUC of the models generated with gradually reduced features on the Wilcoxon test. AUC, area under the curve. *p=0.023; †p=0.009.

CI=0.474–0.719) with the CoxPHR model and 0.843 (95% CI=0.833–0.853) with the GB model (**Fig. 4**). The difference between the AUCs of the second year OS probability was statistically significant according to Hanley and McNeill's method (p=0.0016). We also compared the final results with other machine learning models and logistic regression, and GB model outperformed the others (shown in **Supplementary Table 3**).

5. GB model-guided staging categorizes patients into prognostic subgroups more accurately than the FIGO staging

Based on the survival probability scores predicted by the GB model, patients were divided evenly into 4 subgroups (A, B, C, and D) according to the survival probability scores predicted by the GB model. Each subgroup with survival probability score is shown in **Supplementary Table 4**. Kaplan-Meier curves were drawn by 4 subgroups (A, B, C, and D) according to the

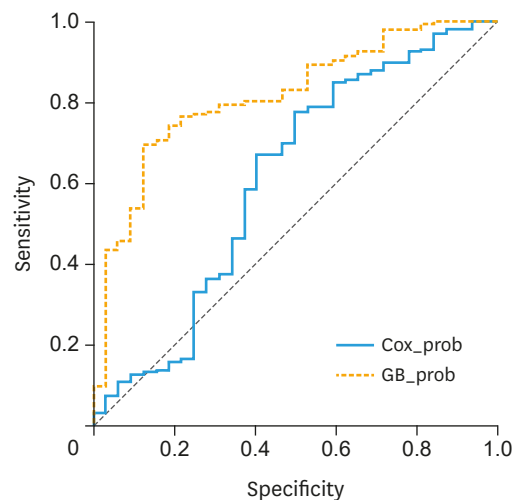


Fig. 4. Comparison of the AUC of the GB model and CoxPHR model in the test set for predicting second year overall survival probability. AUC, area under the curve; GB, gradient boosting; CoxPHR, Cox proportional hazard regression analysis.

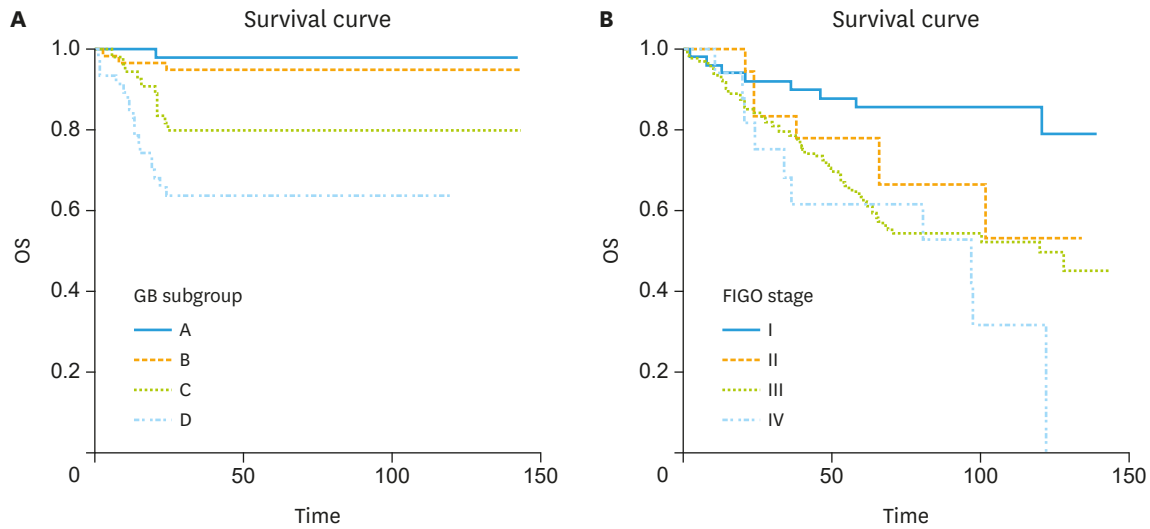


Fig. 5. Kaplan-Meier survival curves for OS in the validation cohort based on the subgroup (A, B, C, and D) according to the second-year OS probability scores predicted by the GB model (A) and subgroup according to FIGO stage (I, II, III, and IV) (B). FIGO, International Federation of Gynecology and Obstetrics; GB, gradient boosting; OS, overall survival.

survival probability scores predicted by the GB model of the test cohort, and Kaplan Meier curves were more discriminately classified (**Fig. 5A**) compared to Kaplan-Meier curves by the FIGO stage groups (I, II, III, and IV; **Fig. 5B**). As result of log-rank tests, the differences between the curves of subgroups A and C were significant ($p=0.003$, hazard ratio [HR]=11.1; 95% CI=3.60–34.57), and the curves of subgroups B and C were significantly different ($p=0.015$; HR=4.25; 95% CI=1.49–12.15). Also, the curves of subgroups C and D were also significantly different ($p=0.0482$; HR=2.1; 95% CI=1.00–4.46). However, the differences between the subgroups by FIGO stages were not clearly shown in the Kaplan-Meier survival curve compared to the GB model (**Fig. 5B**). There were no significant differences between the neighboring FIGO stage curves. There was a significant difference only between the FIGO stage I and IV curves ($p=0.0001$; HR=5.2; 95% CI=1.61–16.52).

DISCUSSION

In this study, we applied a GB model as the machine learning classifier to predict the OS of EOC patients by making use of the sequential CA-125 level during adjuvant chemotherapy after initial debulking surgery and other clinical factors. In the internal validation and external validation, we demonstrated that the GB algorithm accurately predicted the second-year OS rate in EOC patients. Moreover, GB-guided staging identified four distinct subgroups that were classified more clearly than those using FIGO stage.

Accurate prediction for individual patient is still difficult with conventional statistical methods because most of clinical characteristics show multidimensional and non-linear relationship [16]. GB method is one of the machine learning methods that is highly flexible in detecting and recognizing complex non-linear relationships between variables. Compared to FIGO staging predicting with only factors related to tumor, the GB model can provide more individualized prediction based on various factors. To the best of our knowledge, current study is the first study to use a machine learning classifier for the prognostic prediction of endpoint of OS in EOC patients using factors including sequential CA-125 level during

adjuvant treatment after initial debulking surgery. Using our model, clinicians can make more accurate predictions for individual patient.

GB model is capable for recalculating patient's prognosis by adding factors sequentially (i.e., factors indicating the compliance to ongoing therapy, such as CA-125 level). Although CA-125 is a useful marker that is closely related with the progression of disease and the treatment effect in EOC, most published survival prediction models for EOC do not employ CA-125 as a prognostic factor [17-21]. Previous studies have demonstrated the prognostic or predictive values of CA-125 as a single value (e.g., nadirs, thresholds, or normalizations) [11,22,23] or kinetic parameters in different time periods (e.g., percentage decrease) [24,25], half-life [12], or AUC [26]. However, there have been no prediction models for estimating accurate prognosis using CA-125 level, and previous survival-predicting models failed to reflect the result of time sequential CA-125 level. The current study showed that the time sequential serum level of CA-125 can be a useful covariate for predicting OS in EOC patients. This study also showed that GB-based staging more clearly classified the prognostic subgroups compare to FIGO staging.

We noted a few potential limitations. Data were collected retrospectively, and were based on the inherent accuracy of patient records. Therefore, our study may have biases especially with incomplete data collection. For external validation, we used the data of EOC patients from other institution. However, due to the various numbers of variables required, patients without sufficient information were inevitably excluded from the test cohort. These limitations may have led to different characteristics between train and test cohort. There were differences between the training cohort and the external validation cohort except for age, stage, and histology. There were differences in type of treatment between train and test cohort. Although patients with primary debulking surgery and adjuvant chemotherapy were mostly included in both train and test cohort, there were no patients group with neoadjuvant chemotherapy in test cohort. This could lead to differences in survival, and could also affect the accuracy of analysis. However, since it is not comparison study of 2 groups, characteristics of 2 groups are not obligated to be evenly distributed. We think differences between train and test set is allowed, and it may shows generality of the model to work in various patient groups. In development of GB model, second year OS was selected as target for feature selection. It could be limitation of our analysis with expectation of predicting prognosis specifically for only second year OS. Selecting second year OS was not planned based on theoretical background, but performed in the process of obtaining more accurate results. It was selected with highest median AUC of survival prediction among other years, and finally, GB model-guided staging showed more accurate result compared to FIGO staging. Racial and national differences were not considered as the study was performed in the single nation. With regard to the generalizability of this prediction model, validation with other races or other nations may be needed. In our study we used only clinicopathologic information of EOC patients. Lack of genomic data in analysis can be our limitation. Zou et al. [27] reported algorithm of discriminating cancer types for circulating tumor cells or cell-free DNAs in blood trained and validated on a large dataset with high accuracy. In similar perspective, adding genomic data to further analysis could result in higher accuracy of prognostic prediction. With the GB model, additional clinical information would provide more accurate results compared to the currently used Cox proportional hazards regression analysis. The use of larger-scale studies could potentially yield improved prognostic predictions of the GB model in the future.

In summary, our GB model accurately predicted survival outcomes and successfully classified prognostic subgroups in patients with EOC. These may be useful for better estimation of individual outcomes and for selecting better treatment options in EOC patients.

SUPPLEMENTARY MATERIALS

Supplementary Table 1

Clinicopathologic variables used in prediction model development

[Click here to view](#)

Supplementary Table 2

Cox proportional hazard analysis for second-year overall survival in training set

[Click here to view](#)

Supplementary Table 3

Comparison of AUC of the second year overall survival probability between GB model and other machine learning method

[Click here to view](#)

Supplementary Table 4

Subgroups divided by survival probability scores predicted by the GB model

[Click here to view](#)

Supplementary Fig. 1

The prediction accuracy of the gradient boosting model by sequential year interval after primary treatment by comparison of AUCs.

[Click here to view](#)

Supplementary Fig. 2

Area under the curve of the gradient boosting model validated by resampling repetitions with 18 features.

[Click here to view](#)

Supplementary Fig. 3

AUC for predicting second-year overall survival using the Cox proportional hazards model (internal validation) for the training cohort.

[Click here to view](#)

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA Cancer J Clin* 2015;65:5-29.
[PUBMED](#) | [CROSSREF](#)
2. Vaidya AP, Curtin JP. The follow-up of ovarian cancer. *Semin Oncol* 2003;30:401-12.
[PUBMED](#) | [CROSSREF](#)

3. Cotte E, Glehen O, Mohamed F, Lamy F, Falandry C, Golfier F, et al. Cytoreductive surgery and intraperitoneal chemo-hyperthermia for chemo-resistant and recurrent advanced epithelial ovarian cancer: prospective study of 81 patients. *World J Surg* 2007;31:1813-20.
[PUBMED](#) | [CROSSREF](#)
4. Paik ES, Lee YY, Lee EJ, Choi CH, Kim TJ, Lee JW, et al. Survival analysis of revised 2013 FIGO staging classification of epithelial ovarian cancer and comparison with previous FIGO staging classification. *Obstet Gynecol Sci* 2015;58:124-34.
[PUBMED](#) | [CROSSREF](#)
5. Mizuno M, Kajiyama H, Shibata K, Mizuno K, Kawai M, Nagasaka T, et al. Prognostic value of histological type in stage IV ovarian carcinoma: a retrospective analysis of 223 patients. *Br J Cancer* 2015;112:1376-83.
[PUBMED](#) | [CROSSREF](#)
6. Makar AP, Baekelandt M, Tropé CG, Kristensen GB. The prognostic significance of residual disease, FIGO substage, tumor histology, and grade in patients with FIGO stage III ovarian cancer. *Gynecol Oncol* 1995;56:175-80.
[PUBMED](#) | [CROSSREF](#)
7. Mury D, Woelber L, Jung S, Eulenburg C, Choschzick M, Witzel I, et al. Prognostic and predictive relevance of CA-125 at primary surgery of ovarian cancer. *J Cancer Res Clin Oncol* 2011;137:1131-7.
[PUBMED](#) | [CROSSREF](#)
8. Rosendahl M, Høgdall CK, Mosgaard BJ. Restaging and survival analysis of 4036 ovarian cancer patients according to the 2013 FIGO classification for ovarian, fallopian tube, and primary peritoneal cancer. *Int J Gynecol Cancer* 2016;26:680-7.
[PUBMED](#) | [CROSSREF](#)
9. Ayaru L, Ypsilantis PP, Nanapragasam A, Choi RC, Thillanathan A, Min-Ho L, et al. Prediction of outcome in acute lower gastrointestinal bleeding using gradient boosting. *PLoS One* 2015;10:e0132485.
[PUBMED](#) | [CROSSREF](#)
10. Zhang Y, Zhou W, Yuan S, Yuan Q. Seizure detection method based on fractal dimension and gradient boosting. *Epilepsy Behav* 2015;43:30-8.
[PUBMED](#) | [CROSSREF](#)
11. Prat A, Parera M, Peralta S, Perez-Benavente MA, Garcia A, Gil-Moreno A, et al. Nadir CA-125 concentration in the normal range as an independent prognostic factor for optimally treated advanced epithelial ovarian cancer. *Ann Oncol* 2008;19:327-31.
[PUBMED](#) | [CROSSREF](#)
12. Riedinger JM, Wafflart J, Ricolleau G, Eche N, Larbre H, Basuyau JP, et al. CA 125 half-life and CA 125 nadir during induction chemotherapy are independent predictors of epithelial ovarian cancer outcome: results of a French multicentric study. *Ann Oncol* 2006;17:1234-8.
[PUBMED](#) | [CROSSREF](#)
13. Larose DT. *Discovering knowledge in data: an introduction to data mining*. Hoboken, NJ: John Wiley & Sons, Inc.; 2005.
14. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
[PUBMED](#) | [CROSSREF](#)
15. 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.
16. Zhu L, Luo W, Su M, Wei H, Wei J, Zhang X, et al. Comparison between artificial neural network and Cox regression model in predicting the survival rate of gastric cancer patients. *Biomed Rep* 2013;1:757-60.
[PUBMED](#) | [CROSSREF](#)
17. Rutten MJ, Boldingh JH, Schuit E, Trum H, van Driel W, Mol BW, et al. Development and internal validation of a prognostic model for survival after debulking surgery for epithelial ovarian cancer. *Gynecol Oncol* 2014;135:13-8.
[PUBMED](#) | [CROSSREF](#)
18. Barlin JN, Yu C, Hill EK, Zivanovic O, Kolev V, Levine DA, et al. Nomogram for predicting 5-year disease-specific mortality after primary surgery for epithelial ovarian cancer. *Gynecol Oncol* 2012;125:25-30.
[PUBMED](#) | [CROSSREF](#)
19. Teramukai S, Ochiai K, Tada H, Fukushima M; Japan Multinational Trial Organization OC01-01. PIEPOC: a new prognostic index for advanced epithelial ovarian cancer--Japan Multinational Trial Organization OC01-01. *J Clin Oncol* 2007;25:3302-6.
[PUBMED](#) | [CROSSREF](#)
20. Chi DS, Palayekar MJ, Sonoda Y, Abu-Rustum NR, Awtrey CS, Huh J, et al. Nomogram for survival after primary surgery for bulky stage IIIC ovarian carcinoma. *Gynecol Oncol* 2008;108:191-4.
[PUBMED](#) | [CROSSREF](#)

21. Clark TG, Stewart M, Rye T, Smyth JF, Gourley C. Validation of a new prognostic index for advanced epithelial ovarian cancer: results from its application to a UK-based cohort. *J Clin Oncol* 2007;25:5669-70.
[PUBMED](#) | [CROSSREF](#)
22. Tian C, Markman M, Zaino R, Ozols RF, McGuire WP, Muggia FM, et al. CA-125 change after chemotherapy in prediction of treatment outcome among advanced mucinous and clear cell epithelial ovarian cancers: a Gynecologic Oncology Group study. *Cancer* 2009;115:1395-403.
[PUBMED](#) | [CROSSREF](#)
23. Krivak TC, Tian C, Rose GS, Armstrong DK, Maxwell GL. A Gynecologic Oncology Group Study of serum CA-125 levels in patients with stage III optimally debulked ovarian cancer treated with intraperitoneal compared to intravenous chemotherapy: an analysis of patients enrolled in GOG 172. *Gynecol Oncol* 2009;115:81-5.
[PUBMED](#) | [CROSSREF](#)
24. Rustin GJ. Can we now agree to use the same definition to measure response according to CA-125? *J Clin Oncol* 2004;22:4035-6.
[PUBMED](#) | [CROSSREF](#)
25. Lee CK, Friedlander M, Brown C, GebSKI VJ, Georgouloupoulos A, Vergote I, et al. Early decline in cancer antigen 125 as a surrogate for progression-free survival in recurrent ovarian cancer. *J Natl Cancer Inst* 2011;103:1338-42.
[PUBMED](#) | [CROSSREF](#)
26. Mano A, Falcão A, Godinho I, Santos J, Leitão F, Oliveira C, et al. CA-125 AUC as a new prognostic factor for patients with ovarian cancer. *Gynecol Oncol* 2005;97:529-34.
[PUBMED](#) | [CROSSREF](#)
27. Zou J, Wang E. eTumorType, an algorithm of discriminating cancer types for circulating tumor cells or cell-free DNAs in blood. *Genomics Proteomics Bioinformatics* 2017;15:130-40.
[PUBMED](#) | [CROSSREF](#)