Application of Artificial Intelligence for Preoperative Diagnostic and Prognostic Prediction in Epithelial Ovarian Cancer Based on Blood Biomarkers

Eiryo Kawakami[1,2,3], Junya Tabata[2], Nozomu Yanaihara[2], Tetsuo Ishikawa[1,2], Keita Koseki[1], Yasushi Iida[2], Misato Saito[2], Hiromi Komazaki[2], Jason S. Shapiro[2], Chihiro Goto[2], Yuka Akiyama[2], Ryosuke Saito[2], Motoaki Saito[2], Hirokuni Takano[2], Kyosuke Yamada[2], and Aikou Okamoto[2]

Abstract

Purpose: We aimed to develop an ovarian cancer–specific predictive framework for clinical stage, histotype, residual tumor burden, and prognosis using machine learning methods based on multiple biomarkers.

Experimental Design: Overall, 334 patients with epithelial ovarian cancer (EOC) and 101 patients with benign ovarian tumors were randomly assigned to "training" and "test" cohorts. Seven supervised machine learning classifiers, including Gradient Boosting Machine (GBM), Support Vector Machine, Random Forest (RF), Conditional RF (CRF), Naïve Bayes, Neural Network, and Elastic Net, were used to derive diagnostic and prognostic information from 32 parameters commonly available from pretreatment peripheral blood tests and age.

Results: Machine learning techniques were superior to conventional regression-based analyses in predicting multiple clinical parameters pertaining to EOC. Ensemble meth-

ods combining weak decision trees, such as GBM, RF, and CRF, showed the best performance in EOC prediction. The values for the highest accuracy and area under the ROC curve (AUC) for segregating EOC from benign ovarian tumors with RF were 92.4% and 0.968, respectively. The highest accuracy and AUC for predicting clinical stages with RF were 69.0% and 0.760, respectively. High-grade serous and mucinous histotypes of EOC could be preoperatively predicted with RF. An ordinal RF classifier could distinguish complete resection from others. Unsupervised clustering analysis identified subgroups among early-stage EOC patients with significantly worse survival.

Conclusions: Machine learning systems can provide critical diagnostic and prognostic prediction for patients with EOC before initial intervention, and the use of predictive algorithms may facilitate personalized treatment options through pretreatment stratification of patients.

Introduction

Epithelial ovarian cancer (EOC) is classified into at least five distinct histotypes: high-grade serous carcinoma (HGSC), endometrioid carcinoma, clear cell carcinoma, mucinous carcinoma, and low-grade serous carcinoma (LGSC). These histotypes exhibit different morphology, etiology, and biological behavior. According to the World Health Organization (WHO) classification of tumors of the ovary (2014), histotypes are distinguished based on

their histopathologic and immunohistochemical characteristics, as well as the inherent molecular characteristics (1). EOC is surgically and pathologically staged by the International Federation of Gynecology and Obstetrics (FIGO) staging classification, and the current standard of care consists of either primary debulking surgery (PDS) or internal debulking surgery (IDS) following neoadjuvant chemotherapy (NACT; ref. 2). Both histopathology and FIGO staging are considered the gold standard for classification of EOC subgroups and are relevant prognostic factors for stratification (3). Although there is a need for histotype-specific and/or stage-dependent treatment options, most patients with EOC are still treated with a conventional "one-size fits all" approach of surgical intervention and platinum-based combination chemotherapy. Recent clinical application of PARP inhibitors in *BRCA*-deficient ovarian cancers, mostly HGSC, is a major step for individualized cancer treatment strategy that entails genetic testing to define a subgroup of EOC with a specific vulnerability that can be targeted for therapy (4). In addition, the therapeutic benefit of NACT followed by IDS is currently accepted based on large randomized clinical trials in which the prognosis of advanced ovarian cancer treated with NACT plus IDS was not inferior to that treated with PDS followed by chemotherapy (5). However, establishment of patient-selection criteria based on the extent of

[1]Medical Sciences Innovation Hub Program, RIKEN, Yokohama, Kanagawa, Japan. [2]Department of Obstetrics and Gynecology, The Jikei University School of Medicine, Minato-ku, Tokyo, Japan. [3]Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan.

E. Kawakami and J. Tabata contributed equally to this article.

Corresponding Author: Nozomu Yanaihara, The Jikei University School of Medicine, 3-25-8 Nishi-Shinbashi, Minato-ku, Tokyo 105-8461, Japan. Phone: 81-3-3433-1111; Fax: 81-3-3433-1219; E-mail: yanazou@jikei.ac.jp

AACR

Translational Relevance

Identification of variables that predict the patient's characteristics before initial intervention will facilitate selection of more effective therapeutic approaches for epithelial ovarian cancer (EOC). We developed an ovarian cancer–specific prediction approach based on artificial intelligence (AI) using multiple markers in peripheral blood and clinical factors for pretreatment estimation of clinical stages, histotypes, surgical outcomes, and prognosis of patients with EOC. We found that machine learning approach could predict malignant tumors with appreciably high accuracy compared with earlier reports. Moreover, we could show that unsupervised machine learning approach identified subgroups among early-stage EOC patients, which is significantly associated with recurrence-free survival rate. Therefore, this study not only could construct highly accurate predictors of ovarian tumor characteristics but also could propose a usage of AI to reveal difficult-to-recognize clusters of patients from complex combinations of multiple biomarkers. It may be possible to select personalized treatment options by pretreatment stratification of patients with EOC using machine learning–based predictive algorithms.

disease and/or patient condition, as well as universal staging criteria in the NACT setting are thought to be crucial unmet needs to further use this primary treatment modality (5). To select more effective therapeutic approaches for EOC with complex phenotypes, it is important to identify stratification factors that could accurately define patient characteristics before initial intervention. In addition, development of methods to predict treatment outcomes and prognosis is an important paradigm in the realm of personalized medicine (6, 7). Several studies have shown that the diagnostic accuracy can be improved by using a combination of biomarkers and multiple clinical factors (8, 9). Common statistical methods familiar to clinicians are ill-suited for handling complex information; until recently, this has been a major limitation that prevents the extraction of meaningful information from large datasets with multiple input variables. Machine learning is a branch of artificial intelligence (AI) technology that allows computers to "learn" potential patterns from past examples. Use of machine learning approach to predict new data using the identified patterns has helped detect difficult-to-recognize patterns from complex combinations of multiple biomarkers (10).

In this study, we aimed to develop an ovarian cancer–specific prediction approach based on machine learning algorithms using multiple biomarkers and clinical variables for the pretreatment estimation of clinical stages, histotypes, surgical outcomes, and prognosis of patients with EOC.

Materials and Methods

Patients and serum samples

This was a retrospective cohort study of 334 patients with EOC and 101 patients with benign ovarian tumor who were treated between 2010 and 2017 at the Department of Obstetrics and Gynecology, The Jikei University School of Medicine. Tumors were staged in accordance with the FIGO classification (2014). We retrospectively investigated clinicopathologic parameters, including age at diagnosis, clinical stage, residual tumor size after primary surgery, and 32 preoperative peripheral blood biomarkers (Supplementary Table S1). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. The retrospective analysis of clinical information was approved by the ethics committee of The Jikei University School of Medicine [Institutional review Board (IRB) no. 29-138(8754)]. For the study, the IRB issued a waiver for written consent because data collection was retrospective.

Data splitting

Dataset was split into training and test cohorts with repeated random sampling until there was no significant difference (P value ≥ 0.20) between the two cohorts with respect to all variables (Table 1). The P value was calculated using Welch t test for continuous variables and Fisher exact test for categorical variables. This resulted in allocation of 168 patients with EOC and 51 patients with benign ovarian tumor to the training cohort, and 166 patients with EOC and 50 patients with benign ovarian tumor to the test cohort.

Supervised machine learning classifiers

In this study, seven types of supervised machine learning classifiers, including Gradient Boosting Machine (GBM), Support Vector Machine (SVM), Random Forest (RF), Conditional Random Forest (CRF), Naïve Bayes (NB), Neural Network (NN), and Elastic Net (EN), were assessed. We also used logistic regression classifier as the baseline. All classifiers were implemented using R package caret (method "gbm" for GBM, "svmRadial" for SVM, "rf" for RF, "cforest" for CRF, "nb" for NB, "nnet" for NN, and "glmnet" for EN; ref. 11). For ordinal classification, R package ordinalForest was used. Classifiers were trained using repeated 10-fold cross-validation of training dataset, and their predictive performance was evaluated in the test dataset. For calculating variable importance for prediction, 100 sets of independent training were performed using different random seed. The median of variable importance obtained in each training was used as a representative value. Each variable importance is calculated with varImp function of the caret package. In case of ordinal classification, variable importance is provided as varimp using ordfor function of the ordinalForest package.

Confidence of prediction was assessed using Shannon's information gain. When no information about k-class (in which class the patient is included) is available for a patient i, the Shannon's information entropy indicating uncertainty is given by:

$$H(i) = \log_2(k)$$

If a classifier provides prediction probabilities for each class, the entropy will decrease.

$$H_c(i) = \sum_{j=1}^{k} p_j(i)\log_2(p_j(i))$$

Here, $p_j(i)$ is the predicted probability that the patient i is included in class j. By comparing the prior and the posterior entropy, we obtain a measure of total information gain, i.e., information gained by the prediction.

$$IG(i) = H(i) - H_c(i)$$

The individual information gain for each class is given by:

$$IG_j(i) = p_j(i) \times IG(i)$$

Table 1. Clinical characteristics of 435 patients with epithelial ovarian tumor and benign ovarian tumor

	All patients (N = 435)	Training cohort (N = 219)	Test cohort (N = 216)	P value
Age (range)	52.2 (19–87)	51.5 (19–84)	52.9 (23–87)	0.297
Histologic types				
EOC				
High-grade serous (%)	102 (23.4)	46 (21.0)	56 (25.9)	0.258
Endometrioid (%)	66 (15.2)	37 (16.9)	29 (13.4)	0.350
Mucinous (%)	31 (7.1)	18 (8.2)	13 (6.0)	0.457
Clear (%)	98 (22.5)	49 (22.4)	49 (22.7)	1
Others (%)	37 (8.5)	18 (8.2)	19 (8.8)	0.8655
Benign ovarian tumor				
Benign cyst (%)	58 (13.3)	30 (13.7)	28 (13.0)	0.888
Teratoma (%)	43 (9.9)	21 (9.6)	22 (10.2)	0.873
FIGO stage				
I (%)	154 (46.1)	77 (45.8)	77 (46.4)	1
IA (%)	44 (13.2)	21 (12.5)	23 (13.9)	0.748
IC (%)	110 (32.9)	56 (33.3)	54 (32.5)	0.908
II (%)	27 (8.1)	14 (8.3)	13 (7.8)	1
IIA (%)	10 (3.0)	5 (3.0)	5 (3.0)	1
IIB (%)	17 (5.1)	9 (5.4)	8 (4.8)	1
III (%)	128 (38.3)	64 (38.1)	64 (38.6)	1
IIIA (%)	19 (5.7)	10 (6.0)	9 (5.4)	1
IIIB (%)	21 (6.3)	11 (6.5)	10 (6.0)	1
IIIC (%)	88 (26.3)	43 (25.6)	45 (27.1)	0.804
IV (%)	25 (7.5)	13 (7.7)	12 (7.2)	1
IVA (%)	8 (2.4)	3 (1.8)	5 (3.0)	0.501
IVB (%)	17 (5.1)	10 (6.0)	7 (4.2)	0.620
Extent of tumor resection				
Complete (none) (%)	231 (69.2)	120 (71.4)	111 (66.9)	0.407
Optimal (<1 cm) (%)	37 (11.1)	15 (8.9)	22 (13.3)	0.226
Suboptimal (≥1 cm) (%)	66 (19.8)	33 (19.6)	33 (19.9)	1

RF classifier

A RF classifier comprises an ensemble of decision trees (ref. 12; Supplementary Fig. S1A) and is based on two machine learning techniques: bagging and random feature selection. In bagging, each tree is trained using a bootstrap sample of training data. During the process of training, each tree is grown using a particular bootstrap sample. The RF predictive performance during training is assessed using out-of-bag samples, which are not selected in the bootstrap sample. In addition, the RF classifier randomly selects a subset of features in each split node when growing a tree. By virtue of these techniques, the RF classifier avoids overfitting and stratifies samples by considering complex interactions between variables.

Unsupervised RF clustering

An RF dissimilarity measure (13) was used to evaluate the similarity among patients based on multiple variables. The RF dissimilarity was used as input for multidimensional scaling (MDS), which provides a visual representation of the positional relationship among a set of patients. Subsequently, Partitioning Around Medoids (PAM) clustering was applied on the two scaling coordinates of MDS.

Statistical analysis

Correlation between blood markers was evaluated using Spearman rank coefficient. To evaluate the difference in recurrence of cancer, univariate Cox proportional hazards models incorporated in the R package survival were used. Probability values were calculated by the Wald test.

The R codes used in this article are available at https://github.com/eiryo-kawakami/CCR2019_code.

Results

Differentiation of EOC from benign tumor based on multiple preoperative blood markers

To investigate the utility of multiple variables as predictors of ovarian tumor characteristics, we compared multiple logistic regression analysis based on 32 peripheral blood markers to single logistic regression analysis using each marker. Figure 1A shows the ROC curve derived from multiple logistic regression for segregating EOC from benign tumor based on 32 peripheral blood markers in the test cohort (red line). The values for the highest accuracy of the prediction and the area under the ROC curve (AUC) were 86.7% and 0.897, respectively (Supplementary Table S2). These results are superior to those of any single regression, represented by dashed lines (Fig. 1A). When we applied stepwise regression in which the regression model is constructed with a subset of variables, the AUC slightly improved (Fig. 1A, brown line; Supplementary Table S2, 86.7% accuracy and 0.919 AUC). At the same time, the same test dataset with 32 peripheral blood markers was used to predict EOC using several supervised machine learning methods (Fig. 1B; Supplementary Table S2). The highest predictive accuracy and the AUC were 93.7% and 0.976 with GBM, 90.5% and 0.939 with SVM, 92.4% and 0.968 with RF, 93.7% and 0.978 with CRF, 88.6% and 0.954 with NB, 88.0% and 0.883 with NN, and 91.8% and 0.966 with EN, respectively (Supplementary Table S2). Therefore, these supervised machine learning analyses were found to predict more accurately than the conventional multiple logistic regression analysis. It is notable that GBM, RF, and CRF, which are all ensemble methods that combine weak decision trees, displayed the highest performance.

Figure 1.
Differentiation of EOC from benign ovarian tumor based on multiple preoperative blood markers. **A,** ROC curves derived from logistic regression for segregating EOC from benign ovarian tumor. The result of a multiple regression model using all 32 peripheral blood markers is indicated in a red line, whereas single regression results are represented by dashed black lines. **B,** ROC curves for differentiating EOC from benign ovarian tumor using supervised machine learning methods. **C,** Relative importance of variables for segregation of EOC from benign ovarian tumor calculated in the logistic regression and RF. Variable importance is represented as a percentage of the highest value. **D,** Box and jitter plots representing the distribution of top eight important blood markers for distinguishing EOC from benign ovarian tumor. **E,** Correlation between top eight important predictors evaluated using Spearman rank coefficient.
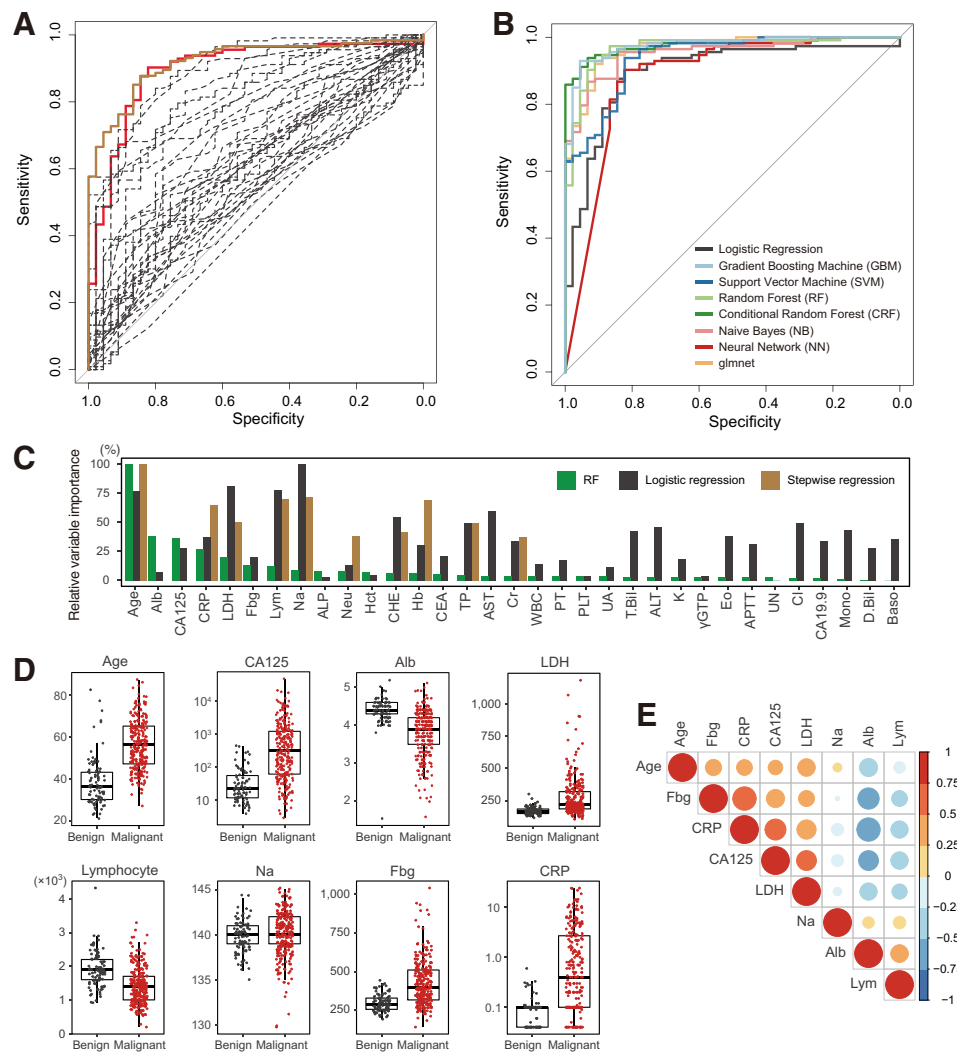
Next, the relative importance of a variable for segregating EOC from benign tumor was calculated with each predictive approach (Fig. 1C). We identified the top eight factors, including age, carbohydrate antigen 125 (CA125), albumin (Alb), lactate dehydrogenase (LDH), lymphocyte, sodium, fibrinogen (Fbg), and C-reactive protein (CRP), as important RF predictors for distinguishing EOC from benign tumor. Standard box plots that present the distribution of each variable between benign and malignant samples are shown in Fig. 1D. In particular, age was a critical variable in all analytical approaches. Interestingly, the importance of any specific variable greatly differed between each predictive method. Logistic regression used various variables including LDH, lymphocyte, and sodium as important predictors, in addition to age (Fig. 1C; Supplementary Table S3). On the contrary, these variables were of less important in the RF that could select subsets of variables to construct weak decision trees and gain accuracy without suffering from overtraining and multicollinearity (14). Although the stepwise regression also selects subsets of variables, it thoroughly relies on the selected variables (Fig. 1C). Use of highly correlated variables in regression analysis is known

to render the model unstable due to multicollinearity. In fact, there were strong positive correlations between Fbg and CRP, CRP and CA125, and CA125 and LDH, and moderate negative correlations between Alb and Fbg, CRP, or CA125 among these variables (Fig. 1E). Accordingly, supervised machine learning algorithms that employ the ensemble method combining weak decision trees such as GBM, RF, and CRF showed the best predictive performance. Therefore, we decided to use RF in subsequent predictive analysis for consistency. Representative classification trees of the RF classifier are presented in Supplementary Fig. S1B and S1C. To evaluate the effect of sample size on machine learning performance, we assessed the highest predictive accuracy and AUC in the RF prediction using different numbers of samples. We randomly selected 20%, 40%, 60%, and 80% of patients from the training and test cohorts. To reduce any potential bias due to random selection, we generated ten independent sets of data for the evaluation. The highest accuracy of prediction and AUC nearly reached saturation when using 60% of patients (Supplementary Fig. S1D and S1E). Thus, based on 32 blood markers, a larger sample size would not provide better predictive performance.
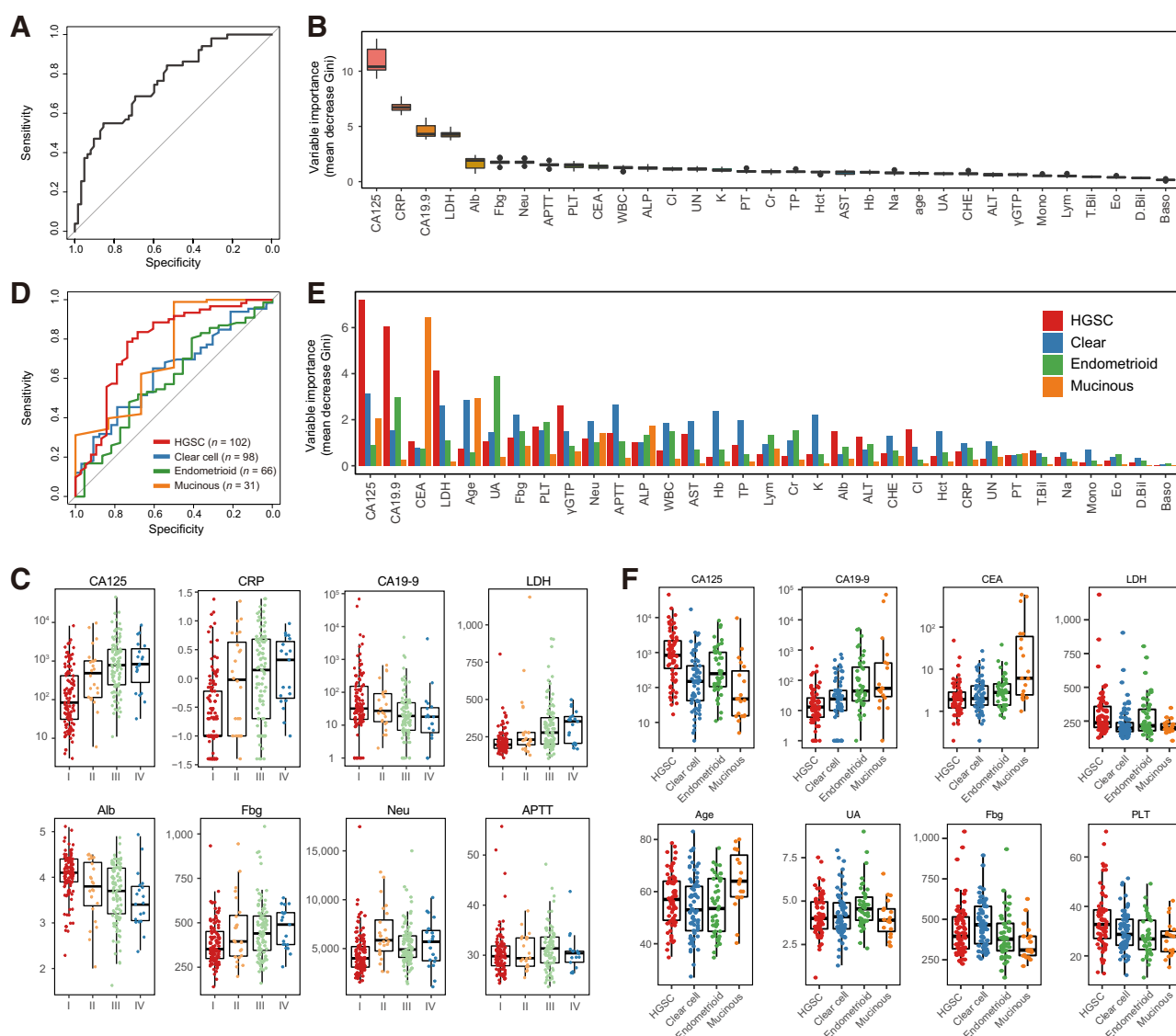
Figure 2.
Prediction of clinical stages and histologic types of EOC with RF classifier. **A** and **D,** ROC curve for RF-based prediction of clinical stages (**A**) and histologic types (**D**) in EOC based on the 32 peripheral blood markers. **B** and **E,** Variable importance for RF-based prediction of clinical stages (**B**) and histologic types (**E**) evaluated as mean decrease in Gini index. The box plot and the bar plot show results from 100 independent training iterations. **C** and **F,** Box and jitter plots representing distribution of top eight important blood markers for RF-based prediction of clinical stages (**C**) and histologic types (**F**).

Prediction of clinical stages and histologic types of EOC with RF classifier

We next attempted to preoperatively predict the clinical stages of EOC, disaggregated into early-stage (stage I/II) and late-stage (stage III/IV), by using the 32 peripheral blood markers with the RF classifier. The values for the highest predictive accuracy and the AUC of the ROC curve were 69.0% and 0.760, respectively (Fig. 2A; Supplementary Table S2). We found CRP and LDH as important parameters for predicting the clinical stage of EOC, in addition to well-known tumor markers such as CA125 and carbohydrate 19-9 (CA19-9), by the mean decrease in Gini index as a measure of variable importance (Fig. 2B). It should be noted that as the clinical stage progressed, CA125, CRP, LDH, Fbg, and

platelet (PLT) increased, whereas other markers including CA19-9 and Alb decreased (Fig. 2C).

We further aimed to evaluate the predictive ability for histologic types of EOC using the same approach. The highest predictive accuracies for high-grade serous, clear cell, endometrioid, and mucinous histotypes were 75.8%, 67.7%, 55.6%, and 96.0%, respectively. The AUC values for the histotypes were 0.785, 0.650, 0.597, and 0.728, respectively (Fig. 2D; Supplementary Table S2). When we evaluated the variable importance for prediction of histotype, CA125 and CA19-9 were the most important predictors for high-grade serous type of EOC (Fig. 2E). As shown in Fig. 2F, these results were consistent with the comparison analyses that showed relatively high CA125 and relatively low CA19-9 in
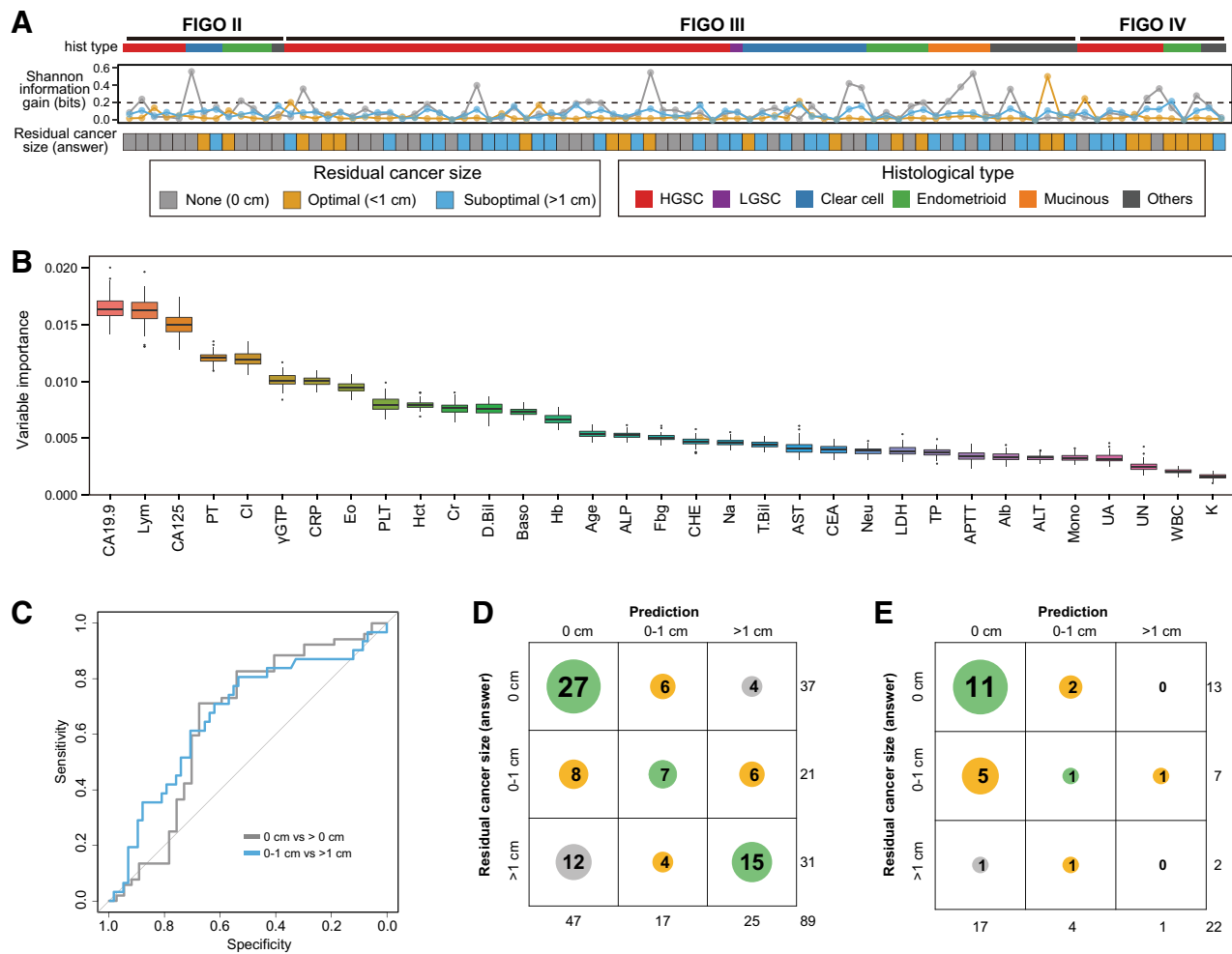
Figure 3.
Prediction of residual tumor size after primary surgery with an ordinal classification method. **A,** Prediction of residual tumor size for individual patients in test cohort. The confidence of prediction for each class is represented as Shannon's information gain. **B,** Variable importance of blood markers for prediction of residual tumor size. Box plot shows results from 100 independent training iterations. **C,** ROC curves for prediction of residual tumor size in patients with EOC. **D** and **E,** Confusion matrix indicating the prediction quality of the RF classification for all predictions (**D**) and for those predictions with high (>0.2 bits) confidence (**E**).

high-grade serous type when compared with the other types of EOC (Fig. 2F). Similarly, we found that CEA, which was an important predictor for mucinous type (Fig. 2E), showed a higher value in mucinous type than in the other types (Fig. 2F).

Prediction of residual tumor with an ordinal classification method

Based on the preoperative blood markers, we also tried to predict residual tumor size. The presence of residual tumor after surgery is a powerful prognostic indicator that affects both progression-free survival (PFS) and overall survival (15). The status of residual tumor is generally classified into three groups based on the extent of resection: "complete" (no residual tumor), "optimal" (<1 cm residual tumor), and "suboptimal" (≥1 cm residual tumor). This classification is an ordinal classification, as the classes exhibit an order (complete < optimal < suboptimal). Standard classification algorithms cannot make use of ordinal information, which impairs the prediction performance. There-

fore, we applied an ordinal classification method on the residual tumor size prediction, which converts the class value into a numeric quantity and applies an RF regression learner to the transformed data (16). Figure 3A shows the prediction results of residual tumor size for individual patients in the test cohort, in which the confidence of prediction for each class is represented as Shannon's information gain. Stage I patients were eliminated from this analysis because there were too few patients with residual tumor. The most important variables for this prediction were CA19-9, lymphocyte, and CA125 (Fig. 3B). The highest predictive accuracy and AUC for distinguishing complete resection (0 cm) from others were 64.9% and 0.697, respectively (Fig. 3C, gray line), whereas those for distinguishing suboptimal resection (≥1 cm) from others were 62.9% and 0.667, respectively (Fig. 3C, light blue line). We identified 16 instances in which the prediction was badly out of order, i.e., misprediction of complete resection as suboptimal or vice versa (Fig. 3D, indicated in gray). Interestingly, predictions designated as high confidence
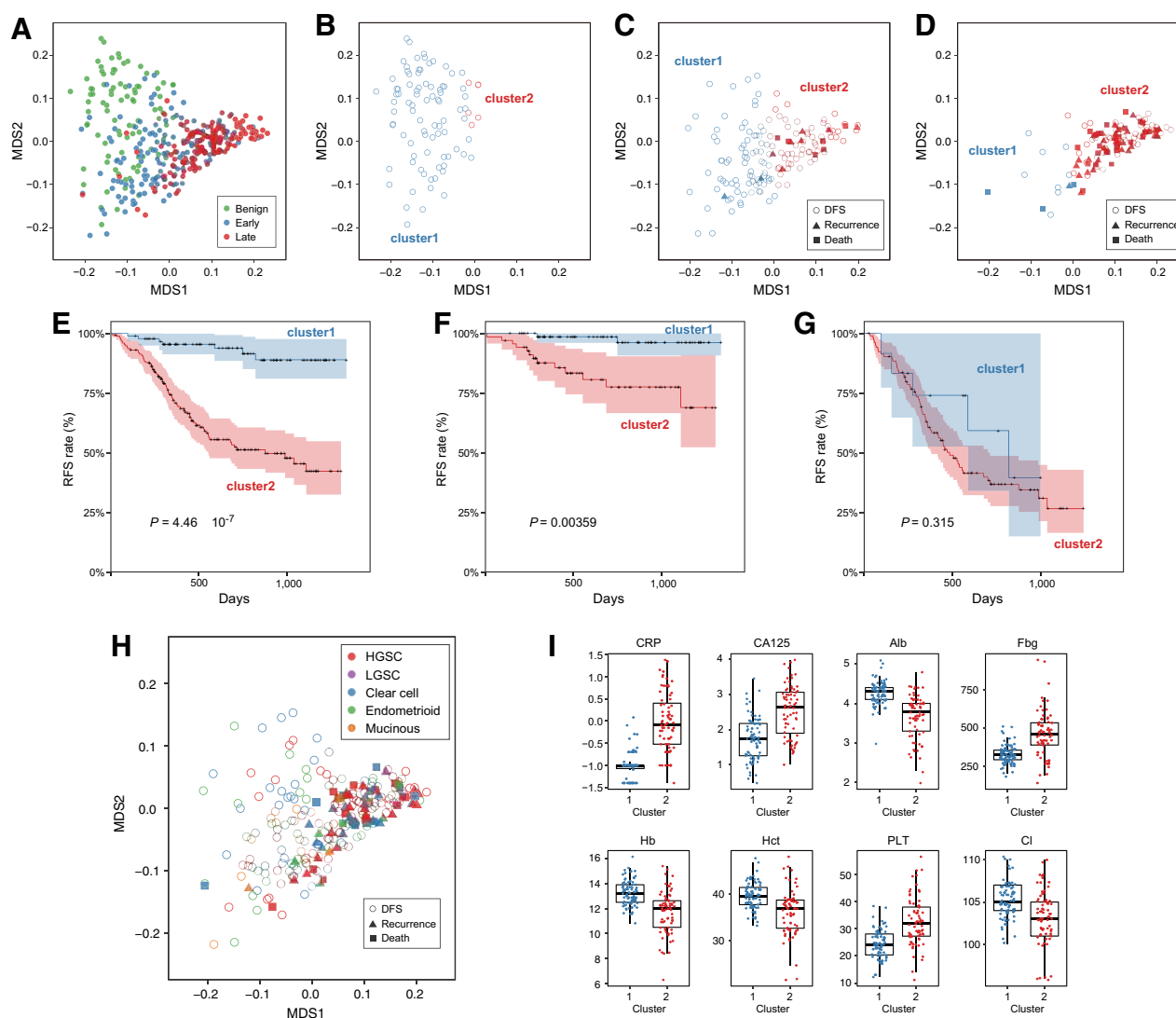
Figure 4.
Unsupervised machine learning clustering associated with prognosis. DFS, disease-free survival. **A,** MDS plot based on the RF dissimilarity analysis for all EOC and benign ovarian tumors. **B–D,** MDS plot for benign ovarian tumors (**B**), early-stage (**C**), and late-stage (**D**) EOC patients clustered into two groups using PAM method. **E–G,** Kaplan–Meier curves indicating RFS of each cluster in all EOC (**E**), early-stage (**F**), and late-stage EOC patients (**G**). *P* values were calculated by Wald test based on univariate Cox proportional hazards models. **H,** MDS plot for all EOC patients indicating histotypes. **I,** Box and jitter plots representing distribution of top eight differential blood markers between the cluster 1 and cluster 2.

(>0.2 bits) contained only 1 such bad misprediction out of the 22 instances (Fig. 3E). Thus, Shannon's information gain can be a useful measure for assessing whether a prediction is wildly out of line in ordinal classification.

Unsupervised clustering analysis using machine learning approach associated with prognosis

Next, unsupervised clustering analysis using an RF dissimilarity measure (13), which can handle mixed variable types and is robust to outliers, was performed to identify specific EOC patient subgroups related to prognosis based on the same 32 preoperative blood markers. MDS plot using the RF dissimilarity as input shows clear separation of benign tumor patients and late-stage EOC patients (Fig. 4A). When PAM clustering method was

applied on the MDS data, almost all benign ovarian tumors were included in cluster 1 (Fig. 4B), whereas most of the late-stage EOC were included in cluster 2 (the area of MDS1 > 0; Fig. 4D). Early-stage EOC were widely distributed among clusters 1 and 2 (Fig. 4C). Examination of the association between the two clusters and clinicopathologic features of EOC showed a statistically significant difference in relapse-free survival (RFS) rate (Fig. 4E, $P = 4.46 \times 10^{-7}$; Table 2). Moreover, we also found a significant difference in RFS rate between the clusters among early-stage EOC patients (Fig. 4F, $P = 0.00359$, Table 2). On the contrary, no significant difference in RFS rate was detected between the two clusters for late-stage EOC (Fig. 4G, $P = 0.315$; Table 2). In addition, we found no clear difference in MDS distribution among EOC histotypes (Fig. 4H). Multiple blood markers

Table 2. Association of RFS with clinicopathologic parameters of patients with EOC

Parameter		HR (95% CI)	P value
All data (N = 334)			
Age	<50 vs. ≥50	1.0 (0.69–1.8)	0.205
Histotype	HGSC vs. others	1.4 (0.95–2.2)	0.0856
Stage	Early vs. late	8.9 (5.1–15)	**9.25 × 10⁻¹⁵**
Residual tumor	0 vs. >0	5.4 (3.6–8.2)	**1.68 × 10⁻¹⁵**
	0–1 vs. >1	4.4 (2.9–6.6)	**2.67 × 10⁻¹²**
Cluster	1 vs. 2	7.4 (3.4–16)	**4.46 × 10⁻⁷**
Early-stage (N = 182)			
Age	<50 vs. ≥50	1.0 (0.96–1)	0.892
Histotype	HGSC vs. others	1.2 × 10⁻⁸ (0–Inf)	0.998
Cluster	1 vs. 2	9.2 (2.1–41)	**0.00359**
Late-stage (N = 152)			
Age	<50 vs. ≥50	1.0 (0.99–1.0)	0.518
Histotype	HGSC vs. others	0.56 (0.36–0.87)	**0.00987**
Residual tumor	0 vs. >0	1.9 (1.2–3.1)	**0.00745**
	0–1 vs. >1	1.8 (1.1–2.7)	**0.0112**
Cluster	1 vs. 2	1.6 (0.64–4)	0.315

NOTE: There were too few early-stage EOC patients with residual tumor. A definition for the significance of bold is P value of < 0.05.
Abbreviations: CI, confidence interval; HR, hazard ratio; Inf, infinity.

including CRP, CA125, Alb, Fbg, hemoglobin (Hb), Hct, PLT, and chloride were significantly different between the early-stage EOC in the two clusters (Fig. 4I). The univariate Cox proportional hazards model based on each clinicopathologic parameter showed that stage and residual tumor size were also significant prognostic factors in all EOC (Table 2).

Discussion

Use of machine learning algorithms based on AI technology for diagnostic and prognostic assessment has been widely accepted in the context of some cancers (6, 9). It is clear that this innovative approach is an important tool in the realm of precision medicine that may facilitate the selection of optimal treatment strategies. In addition, the ability of AI models to discover embedded patterns within data by handling numerous factors at once may lead to a better understanding of the complex mechanisms that underlie carcinogenesis and cancer progression. However, the machine learning algorithm that provides the greatest diagnostic and prognostic power for a given set of variables is poorly understood. Our approach allowed for the comparison of multiple supervised learning algorithms to identify the approach with the most favorable performance. Ovarian cancer comprises multiple heterogeneous features containing various clinical stages and several histopathologies with varying grades. The current standard of treatment, with its "one-size fits all" approach, is no longer a sufficient strategy in light of the recent development and evaluation of targeted therapies and our growing knowledge of the molecular mechanisms of this disease. Currently, the lack of ability to accurately identify clinically meaningful patient subsets before initial treatment has been a key limitation in clinical settings. Therefore, predicting clinical characteristics of EOC based on preoperative information and stratification of patients with respect to prognosis is a fundamental approach toward individualized optimal medical care. In a study, preoperative monocyte-to-lymphocyte ratio in peripheral blood of patients with ovarian cancer was identified as a predictor of clinical characteristics based on binary logistic regression analysis (17). In a recent study, AI systems were used for prognostic assessment of patients with

ovarian cancer based on basic clinical variables including age, FIGO stage, histopathology with tumor grade, and CA125 (18). In this study, we investigated the ability of multiple machine learning methods to predict the basic characteristics of patients with EOC based on readily available biomarkers. We found that ensemble classifiers such as RF that incorporate weak decision trees were able to preoperatively predict various clinical variables such as stages and histotypes (high-grade serous and mucinous) of EOC with appreciable accuracy (69.0% accuracy and 0.760 AUC for clinical stages; 75.8% accuracy and 0.785 AUC for high-grade serous; 96.0% accuracy and 0.728 AUC for mucinous). The underperformance of these classifiers with regard to clear cell (67.7% accuracy and 0.650 AUC) and endometrioid histotypes (55.6% accuracy and 0.597 AUC) may result from the lack of particularly strong distinguishing characteristics of these tumors at the level of serum biomarkers. Nevertheless, these results indicate that AI technology may provide valuable diagnostic information based on preoperative biomarkers, which may facilitate a personalized treatment strategy before the primary therapeutic approach in EOC. In addition, based on the thorough comparison of different variables using supervised machine learning techniques, this study may provide valuable information to clinicians regarding variables that are the most useful for patient stratification.

Identification of reliable biomarkers that are able to predict surgical outcomes in advance would facilitate the identification of patients with advanced EOC who may benefit from PDS (19). It is well accepted that residual disease following upfront surgery strongly correlates with patient survival and that complete gross cytoreduction to no residual disease status appears to be associated with the best overall outcomes (15). However, patients with advanced stage EOC who are preoperatively predicted to have residual disease based on our machine learning approach may be superior candidates for NACT. Here, we report the use of ordinal classification method to predict surgical outcomes in terms of residual tumor size in stage II–IV EOC patients with a 64.9% accuracy and AUC of 0.697 (0 cm vs. >0 cm) based solely on preoperative information. Recently, large transcriptional profiles of primary debulked EOC tumors have been used to identify genomic signatures that had the potential to accurately predict suboptimal cytoreduction as the outcome of PDS (20). The caveat to this approach is that surgery is needed to obtain the samples for analysis, at which point the outcome of cytoreduction would already be known. Although not assessed in this study, preoperative prediction of tumor chemosensitivity may have a profound impact on treatment decision-making vis-a-vis initiation of NACT; therefore, further efforts should be made to establish methods for predicting tumor chemosensitivity.

In a previous study, a combination of serum tumor markers and age with or without ultrasound findings was used to predict ovarian cancer in patients with adnexal masses (21). In particular, multivariate logistic regression analysis was used to differentiate stage I EOC from benign ovarian tumor using HE4, CA125, CEA, and patient's age (AUC: 0.797; ref. 8). It has been recently reported that preoperative serum CRP levels could be of additional value to CA125 in the differential diagnosis of ovarian tumor (22). In our study, segregation of EOC from benign ovarian tumor was achieved with a high accuracy (~94%; AUC: ~0.98) by several supervised machine learning approaches, which clearly outperformed standard regression analysis and the existing prediction models. Furthermore, factors such as Alb, LDH, lymphocyte,

sodium, and Fbg were found to be useful in differentiating EOC from benign ovarian tumor, in addition to the known tumor markers; these findings suggest that the supervised machine learning analysis can help identify new biomarkers that are not identified by conventional multiple regression analysis. However, as shown in Fig. 1E, there were strong correlations among some of these important explanatory variables. It is well-known that multicollinearity among explanatory variables may pose a problem with increase in the number of variables. In this context, multicollinearity among these explanatory variables could be excluded by using the ensemble method that incorporates weak decision trees including RF.

As the approach used in this study did not consider any information from imaging studies or pretreatment biopsies, the ability to accurately predict the clinical behavior and treatment outcome before intervention was limited. However, manipulation of large datasets from high-throughput sequencing analysis such as RNA sequencing of preoperative peripheral blood may improve prediction performance. Therefore, further validation efforts should be made by increasing the number of input variables based on the machine learning approach robust to overfitting in a larger independent cohort. As tumors grow over time, signaling between the tumor and its niche, consisting of fibroblasts, infiltrating immune cells, and endothelial cells, also evolves. It is believed that chemoresistant and highly aggressive tumors become so, in part, due to permissive signals that originate in the niche (23). Despite the importance of the tumor environment, clinicians still rely nearly exclusively on tumor-specific markers for prognostic assessment and treatment decision-making. Changes in parameters obtained from preoperative peripheral blood investigations are inherently a combination of tumor-specific and niche-specific factors. The machine learning approach used in this study identified systemic factors such as Alb, LDH, lymphocyte, and sodium as important factors in malignancy; this approach may identify patients with protumor niches, which may significantly influence the choice and timing of treatment.

Accurate prognostic prediction tools aid clinical decision-making for the management of EOC. The supervised machine learning approach in this study revealed the association of preoperative blood markers with important features of EOC, which may be used for stratification of patients. This prompted us to investigate the direct correlation of these makers with prognosis of EOC patients. Unsupervised clustering analysis based on 32 preoperative blood markers was able to segregate EOC subgroups which were clearly associated with clinical stage and prognosis. Importantly, the series of unsupervised machine learning approach revealed two clusters in early-stage EOC associated with prognosis, which could be classifiable preoperatively. In a previous study, readily available biomarkers in clinical settings including indicators of the systemic inflammatory response and pretherapeutic coagulation-related factors were shown to be of prognostic relevance in patients with EOC (24). A recent meta-analysis of data from 13 studies

($n = 3,467$) showed that both high neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio are associated with unfavorable prognosis of patients with EOC (25). In addition, elevated levels of pretreatment plasma D-dimer, Fbg, and PLT were found useful in predicting disease progression and survival outcomes of patients with EOC (26–28). These reports support our findings; indeed, additional studies using independent datasets are required to investigate how we can utilize this preoperative blood signature for accurate prognostic assessment of patients with EOC. Furthermore, future studies should investigate the use of AI-based machine learning algorithms to identify predictive features in time series of preoperative blood values, which might significantly expand the accuracy of prognosis and warrants further investigation.

In conclusion, this study demonstrates that AI-based algorithms are powerful tools that may provide critical information for diagnostic and prognostic assessment of patients with EOC before initial intervention.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: E. Kawakami, J. Tabata, N. Yanaihara, Y. Iida, H. Komazaki, M. Saito, K. Yamada
Development of methodology: E. Kawakami, J. Tabata, K. Koseki, Y. Iida
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Tabata, Y. Iida, M. Saito, H. Komazaki, C. Goto, Y. Akiyama, R. Saito, H. Takano
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): E. Kawakami, J. Tabata, N. Yanaihara, T. Ishikawa, K. Koseki, K. Yamada
Writing, review, and/or revision of the manuscript: E. Kawakami, J. Tabata, N. Yanaihara, T. Ishikawa, J.S. Shapiro, R. Saito
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): E. Kawakami, J. Tabata, N. Yanaihara, M. Saito, C. Goto, Y. Akiyama, R. Saito
Study supervision: E. Kawakami, J. Tabata, M. Saito, A. Okamoto

Acknowledgments

References

1. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO classification of tumors of female reproductive organs. Lyon: International Agency for Research on Cancer;2014.

2. Vergote I, Tropé CG, Amant F, Kristensen GB, Ehlen T, Johnson N, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. N Engl J Med 2010;363:943–53.

3. Heintz AP, Odicino F, Maisonneuve P, Quinn MA, Benedet JL, Creasman WT, et al. Carcinoma of the ovary. Int J Gynecol Obstet 2006;95:S161–92.

4. Grunewald T, Lederman JA. Targeted therapies for ovarian cancer. Best Pract Res Clin Obstet Gynaecol 2017;41:139–52.

5. McGee J, Bookman M, Harter P, Marth C, McNeish I, Moore KN, et al. Fifth ovarian cancer consensus conference: individualized therapy and patient factors. Ann Oncol 2017;28:702–10.

6. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8–17.

7. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5:845–56.

8. Kondalsamy-Chennakesavan S, Hackethal A, Bowtell D, Obermair A. Differentiating stage 1 epithelial ovarian cancer from benign ovarian tumours using a combination of tumour markers HE4, CA125, and CEA and patient's age. Gynecol Oncol 2013;129:467–71.

9. Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. Breast Cancer Res Treat 2010;120:293–308.

10. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform 2006;2:59–77.

11. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26.

12. Breiman L. Random forests. Mach Learn 2001;45:5–32.

13. Shi T, Horvath S. Unsupervised learning with random forest predictors. J Comput Graph Stat 2006;15:118–38.

14. Kleinberg E. An overtraining-resistant stochastic modeling method for pattern recognition. Ann Stat 1996;24:2319–49.

15. I Du Bois A, Reuss A, PujadeLauraine E, Harter P, RayCoquard I, Pfisterer J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the arbeitsgemeinschaft gynaekologische onkologie studiengruppe ovarialkarzinom (AGO-OVAR) and the groupe d'Investigateurs nationaux pour les etudes des cancers de l'Ovaire (GINECO). Cancer 2009;115:1234–44.

16. Frank E, Hall M. A simple approach to ordinal classification. In:De Raedt L, Flach P, editors. European Conference on Machine Learning. Heidelberg, Berlin: Springer; 2001. p. 145–56.

17. Xiang J, Zhou L, Li X, Bao W, Chen T, Xi X, et al. Preoperative monocyte-to-lymphocyte ratio in peripheral blood predicts stages, metastasis, and histological grades in patients with ovarian cancer. Transl Oncol 2017;10:33–9.

18. Enshaei A, Robson CN, Edmondson RJ. Artificial intelligence systems as prognostic and predictive tools in ovarian cancer. Ann Surg Oncol 2015;22:3970–5.

19. Wei W, Giulia F, Luffer S, Kumar R, Wu B, Tavallai M, et al. How can molecular abnormalities influence our clinical approach. Ann Oncol 2017;28:viii16–24.

20. Riester M, Wei W, Waldron L, Culhane AC, Trippa L, Oliva E, et al. Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples. J Natl Cancer Inst 2014;106:1–12.

21. Håkansson F, Høgdall EV, Nedergaard L, Lundvall L, Engelholm SA, Pedersen AT, et al. Risk of malignancy index used as a diagnostic tool in a tertiary centre for patients with a pelvic mass. Acta Obstet Gynecol Scand 2012;91:496–502.

22. Reiser E, Aust S, Seebacher V, Reinthaller A, Helmy-Bader S, Schwameis R, et al. Preoperative C-reactive protein serum levels as a predictive diagnostic marker in patients with adnexal masses. Gynecol Oncol 2017;147:690–4.

23. Prieto-Vila M, Takahashi RU, Usuba W, Kohama I, Ochiya T. Drug resistance driven by cancer stem cells and their niche. Int J Mol Sci 2017;18:2574.

24. Wang X, Wang E, Kavanagh JJ, Freedman RS. Ovarian cancer, the coagulation pathway, and inflammation. J Transl Med 2005;3:1–20.

25. Zhao Z, Zhao X, Lu J, Xue J, Liu P, Mao H. Prognostic roles of neutrophil to lymphocyte ratio and platelet to lymphocyte ratio in ovarian cancer: a meta-analysis of retrospective studies. Arch Gynecol Obstet 2018;297:849–57.

26. Luo Y, Kim HS, Kim M, Lee M, Song YS. Elevated plasma fibrinogen levels and prognosis of epithelial ovarian cancer: a cohort study and meta-analysis. J Gynecol Oncol 2017;28:1–12.

27. Man YN, Wang YN, Hao J, Liu X, Liu C, Zhu C, et al. Pretreatment plasma D-dimer, fibrinogen, and platelet levels significantly impact prognosis in patients with epithelial ovarian cancer independently of venous thromboembolism. Int J Gynecol Cancer 2015;25:24–32.

28. Allensworth SK, Langstraat CL, Martin JR, Lemens MA, McGree ME, Weaver AL, et al. Evaluating the prognostic significance of preoperative thrombocytosis in epithelial ovarian cancer. Gynecol Oncol 2013;130:499–504.