International Journal of Women's Health

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Random Survival Forest Model for Predicting Recurrence of Postoperative Ovarian Endometrioma

Yayun Zhang[1], Meng Liu[1], Zihan Zhao[2], Yan Zhang[1], Shunyu Hou[1], Aifen Wang[3]

[1]Department of Gynecology, Suzhou Municipal Hospital Affiliated to Nanjing Medical University, Suzhou, Jiangsu, 215001, People's Republic of China; [2]Department of Epidemiology, George Washington University Milken Institute School of Public Health, Washington, DC, 20037, USA; [3]Department of Gynecology, Jiangnan University Medical Center, Wuxi, Jiangsu, 214002, People's Republic of China

Correspondence: Aifen Wang, Department of Gynecology, Jiangnan University Medical Center, 68 Zhongshan Road, Liangxi Strict, Wuxi, Jiangsu, 214002, People's Republic of China, Tel +86 512 62362003, Email aifen.wang@njmu.edu.cn

Objective: To develop and compare a random survival forest (RSF) model and a Cox regression model—for predicting postoperative recurrence in patients with ovarian endometrioma.

Methods: The primary outcomes of the study were the recurrence of ovarian endometrioma and the time to recurrence. Recurrence-related clinical variables were assessed using two methods: Cox regression analysis and the random survival forest algorithm. The performance of the prediction models was evaluated using time dependent Receiver Operating Characteristic (time-dependent ROC) curve, calibration curve, and Decision Curve Analysis (DCA) curve.

Results: 109 patients experienced recurrence of postoperative ovarian endometrioma, while 469 patients did not experience recurrence. The multivariate Cox regression model identified four independent variables. The RSF model was constructed using 12 important clinical variables. In the training set, the Cox model presented a lower C-index compared to the RSF model (0.699 vs. 0.987). However, in the testing set, the Cox model had a slightly higher C-index compared to the RSF model (0.771 vs. 0.763). The classification accuracy of the RSF model and the traditional Cox regression model was evaluated using two performance metrics: NRI and IDI. The NRI at 3 years and 5 years were 0.728 and 0.775 respectively. The IDI at 3 years and 5 years were 0.599 and 0.603 respectively.

Conclusion: The predictive performance of the RSF model was slightly better than that of the traditional Cox model for predicting postoperative recurrence in patients with ovarian endometrioma.

Keywords: ovarian endometrioma, random survival forest, cox model, prognosis

Introduction

Endometriosis is characterized by chronic inflammation resulting from the presence of endometrial stroma and glandular tissue outside the uterus. Among the various types of endometriosis, ovarian endometriosis is the most prevalent and often requires surgical intervention.[1,2] However, one of the critical challenges in managing ovarian endometriosis is its high postoperative recurrence rate. Factors such as removal of endometrioma larger than 8 cm, younger age (less than 25 years old), and preoperative cyst rupture were identified as risks for recurrence of ovarian endometrioma.[3] Currently, there is a lack of predictive model associated with the recurrence of ovarian endometriosis, posing difficulties in accurately predicting the risk of postoperative recurrence in affected patients. In this study, a retrospective analysis of clinical variables was conducted and subsequently constructed two models: the Cox regression model and the random survival forest (RSF) model. These models were applied to analyze prognostic factors for patients, aiming to provide medical evidence for personalized prognosis evaluation and clinical decision-making guidance. RSF offers several advantages over Cox proportional hazards models. Firstly, RSF does not rely on specific assumptions about the underlying distribution of survival times, making it more flexible. Unlike Cox models, which assume constant hazard ratios

https://doi.org/10.2147/IJWH.S578547
Received: 8 December 2025
Accepted: 3 March 2026
Published: 11 March 2026

over time, RSF can handle non-proportional hazards and capture non-linear relationships between predictors and survival outcomes without explicit specification. This enables more accurate modeling of complex data patterns. Additionally, RSF effectively handles missing data without requiring imputation techniques. Moreover, RSF provides a measure of variable importance, aiding in the identification of the most influential predictors in the model, thus enhancing understanding of the underlying mechanisms driving survival outcomes. Furthermore, RSF incorporates the advantages of the general random forest algorithm and mitigates overfitting issues through two random sampling processes.

Materials and Methods

Research Participants

Clinical data were collected retrospectively from patients who underwent resection of ovarian endometrioma at our hospital between March 2013 and December 2018. The inclusion criteria were as follows: patients who met the diagnostic criteria for ovarian endometrioma, were aged < 45 years, underwent resection of ovarian endometrioma, and received postoperative pathological diagnosis confirming ovarian endometrioma. Patients with a previous history of ovarian endometrioma, preoperative cyst rupture, history of hysterectomy, or those with abnormal function of important organs or malignant tumors were excluded. The recurrence of cysts after surgery was monitored, with a mean follow-up period of 90 months (ranging from 60 months to 129 months).

Statistical Methods

All data analysis was conducted using R 4.3.0 software. The "caret" package was utilized to randomly split the dataset into a training set and a test set, maintaining a 7:3 ratio. Subsequently, the "Compare Groups" package was applied to conduct statistical tests to evaluate whether the two groups were balanced in terms of various variables. Continuous data were summarized using mean and standard deviations (SD). Comparisons between two groups were performed using independent sample t-tests or Mann–Whitney *U*-tests, depending on the distribution of the data. Categorical data were described using counts (n) and percentages (%). The $\chi 2$ (chi-square) test or Fisher's exact test was applied to compare categorical data between groups, depending on the sample size and expected cell frequencies.

In the training set, the Cox regression model was constructed as follows: The "survival" package in R was utilized to conduct univariate Cox analysis. Variables with p-values less than 0.1 in the univariate Cox analysis were selected for inclusion in the multivariate Cox regression model. A multivariate Cox regression model was established using the selected variables. Finally, a nomogram model was constructed using the "rms" package in R to visualize the results of the Cox regression analysis.

For the random forest model construction: Variable screening was performed in the training set using the Variable Importance in Projection (VIMP) method and the minimum depth method available in the "randomForestSRC" package.[4] Important variables identified through the screening process were incorporated into the random survival model modeling. The "mlr3" package was applied for hyperparameter tuning and ten-fold cross-validation of the random survival forest model to optimize its performance and obtain the final model.

Both models were evaluated their performance using various metrics, including time-dependent ROC curve, area under the ROC curve (AUC), calibration curve, and Decision Curve Analysis (DCA). Additionally, the Bootstrap method with 500 resampling iterations was employed to calculate the DCA. Finally, the classification accuracy of the two models was measured by the Integrated Discrimination Improvement Index (IDI) and the Net Reclassification Index (NRI). $P < 0.05$ was defined as statistical significance.

Results

Variables Were Screened by Univariate Cox Analysis

Initially, a total of 628 patients with resection of ovarian endometrioma were screened. Ultimately, the study recruited 578 patients with ovarian endometrioma, of whom 109 experienced recurrent cases. Baseline clinical variables were compared between the training set and the testing set. No statistically significant differences were observed between the two sets (Table 1). Postoperative recurrence (recurrence = 1, non-recurrence = 0) and

Table 1 Characteristics of Clinical Variables

Variable	Training Set N = 434	Testing Set N = 144	P value
Age at surgery, mean (SD), year	32.50 (5.71)	32.60 (5.61)	0.875
Recurrence status, N (%)			0.252
No recurrence	347 (80.00%)	122 (84.70%)	
Recurrence	87 (20.00%)	22 (15.30%)	
Age of menarche, mean (SD), year	14.10 (0.85)	14.10 (0.86)	0.448
Duration of menstruation, mean (SD), year	5.93 (1.40)	5.84 (1.39)	0.488
Cycles of menstruation, mean (SD), year	29.30 (3.85)	29.20 (3.81)	0.747
Volumes of menstruation, N (%)			0.470
<20mL	4 (0.92%)	3 (2.08%)	
20mL to 80mL	421 (97.0%)	139 (96.5%)	
>80mL	9 (2.07%)	2 (1.39%)	
Dysmenorrhea VRS-5, median (IQR)	0 (0; 1)	0.00 (0; 1)	0.769
Preoperative number of deliveries, mean (SD)	0.65 (0.58)	0.59 (0.56)	0.275
Preoperative number of abortions, mean (SD)	0.64 (0.91)	0.56 (0.88)	0.348
Number of deliveries after surgery, mean (SD)	0.42 (0.61)	0.50 (0.61)	0.160
Number of abortions after surgery, mean (SD)	0.11 (0.37)	0.08 (0.30)	0.378
Family history of endometriosis, N (%)			1.000
No	425 (97.90%)	142 (98.60%)	
Yes	9 (2.07%)	2 (1.39%)	
Hypertension, N (%)			1.000
No	417 (96.10%)	139 (96.50%)	
Yes	17 (3.92%)	5 (3.47%)	
Management with GnRH-a, N (%)			0.985
No	138 (31.80%)	45 (31.20%)	
Yes	296 (68.20%)	99 (68.80%)	
Diameter of cyst, mean (SD), mm	54.30 (17.70)	53.50 (17.90)	0.655
Number of cysts, N (%)	1.52 (0.82)	1.53 (0.77)	0.877
Unilateral or bilateral cyst, N (%)			0.892
Unilateral	288 (66.40%)	94 (65.30%)	
Bilateral	146 (33.60%)	50 (34.70%)	

(*Continued*)

Table 1 (Continued).

Variable	Training Set N = 434	Testing Set N = 144	P value
Coexisted with adenomyosis, N (%)			0.467
No	400 (92.20%)	136 (94.40%)	
Yes	34 (7.83%)	8 (5.56%)	
Complicated with deep endometriosis, N (%)			1.000
No	426 (98.20%)	142 (98.60%)	
Yes	8 (1.84%)	2 (1.39%)	
Adhesion of pouch of Douglas, N (%)			0.387
No	200 (46.10%)	73 (50.70%)	
Yes	234 (53.90%)	71 (49.30%)	
Adhesion of cyst, N (%)			0.542
No	25 (5.76%)	11 (7.64%)	
Yes	409 (94.20%)	133 (92.40%)	
Surgical approach, N (%)			0.753
Laparotomy	25 (5.76%)	10 (6.94%)	
Laparoscopy	409 (94.20%)	134 (93.10%)	
Suture of cyst, N (%)			0.580
No	261 (60.10%)	91 (63.20%)	
Yes	173 (39.90%)	53 (36.80%)	
rAFS staging, N (%)			0.413
3	180 (41.50%)	66 (45.80%)	
4	254 (58.50%)	78 (54.20%)	
Endometriosis invasion out of ovaries, N (%)			1.000
No	249 (57.40%)	82 (56.90%)	
Yes	185 (42.60%)	62 (43.10%)	
CA125, mean (SD), (U/mL)	86.80 (116.00)	83.30 (102)	0.731
CA199, mean (SD), (U/mL)	59.30 (126.00)	49.70 (61.00)	0.223
Fasting blood glucose, mean (SD), (mmol/l)	4.75 (0.82)	4.81 (1.02)	0.489
Triglyceride, mean (SD), (mmol/l)	0.90 (0.57)	0.85 (0.57)	0.364
Cholesterol, mean (SD), (mmol/l)	4.17 (0.78)	4.13 (0.75)	0.598
CRP, mean (SD), (mg/l)	3.59 (15.40)	3.27 (10.20)	0.782
White cell counts, mean (SD), (*10^9/l)	5.97 (1.78)	6.01 (1.81)	0.841
Recurrence time, mean (SD), month	81.10 (28.10)	80.50 (26.40)	0.813

(*Continued*)

Table 1 (Continued).

Variable	Training Set N = 434	Testing Set N = 144	P value
Blood types, N (%)			0.541
A	135 (31.10%)	54 (37.50%)	
B	130 (30.00%)	39 (27.10%)	
AB	124 (28.60%)	36 (25.00%)	
O	45 (10.40%)	15 (10.40%)	

Abbreviations: SD, Standard deviation; rAFS: IQR, Interquartile range; rAFS, Revised American Fertility Society; CA125, Carbohydrate antigen 125; CA199, Carbohydrate antigen 199, CRP, C-reactive protein; VRS, Verbal Rating Scale-5.

recurrence time were set as the primary outcomes. Dysmenorrhea VRS-5 (Verbal Rating Scale-5), diameter of ovarian cyst, age at the time of surgery, number of preoperative abortions, days of menstruation, and intraoperative revised American Fertility Society (rAFS) staging were found to be associated with postoperative recurrence of ovarian endometrioma based on univariate Cox analysis. These associations were statistically significant ($p < 0.05$) (Table 2).

Table 2 Univariate Cox Analysis of Clinical Characteristics

	Variable	β	HR (95% CI)	P value
General information	Age at surgery	−0.055	0.946 (0.910 to 0.984)	<0.001
	Age of menarche	0.123	1.130 (0.883 to 1.450)	0.329
	Cycles of menstruation	0.010	1.010 (0.956 to 1.070)	0.727
	Duration of menstruation	0.162	1.180 (1.010 to 1.370)	0.037
	Volumes of menstruation	−0.522	0.594 (0.173 to 2.040)	0.408
	Dysmenorrhea VRS-5	0.283	1.330 (1.130 to 1.560)	< 0.001
	Preoperative number of deliveries	−0.208	0.812 (0.559 to 1.180)	0.274
	Preoperative number of abortions	−0.408	0.665 (0.493 to 0.897)	< 0.001
	Hypertension	−0.092	0.912 (0.288 to 2.890)	0.876
	Blood types	−0.399	0.671 (0.385 to 1.170)	0.158
	Family history of endometriosis	0.703	2.020 (0.740 to 5.520)	0.170
	White cell counts	−0.038	0.963 (0.854 to 1.090)	0.539
	CA125	−0.001	0.999 (0.997 to 1.000)	0.414
	CA199	−0.002	0.998 (0.996 to 1.000)	0.267
	CRP	−0.011	0.989 (0.966 to 1.010)	0.391
	Fasting blood glucose	−0.158	0.854 (0.607 to 1.200)	0.364
	Triglyceride	−0.262	0.769 (0.495 to 1.200)	0.245
	Cholesterol	−0.051	0.950 (0.727 to 1.240)	0.706

(*Continued*)

Table 2 (Continued).

	Variable	β	HR (95% CI)	P value
Intraoperative variables	Surgical approach	0.663	1.940 (0.613 to 6.150)	0.259
	Diameter of cyst	0.016	1.020 (1.000 to 1.030)	< 0.001
	Number of cysts	−0.073	0.93 (0.710 to 1.220)	0.598
	Adhesion of cyst	1.040	2.840 (0.698 to 11.500)	0.145
	Unilateral or bilateral cyst	−0.087	0.917 (0.584 to 1.440)	0.705
	Suture of cyst	0.070	1.070 (0.699 to 1.650)	0.748
	Coexisted with adenomyosis	0.353	1.420 (0.713 to 2.840)	0.317
	Complicated with deep endometriosis	0.237	1.270 (0.312 to 5.150)	0.741
	Adhesion of pouch of Douglas	0.207	1.230 (0.802 to 1.880)	0.342
	Endometriosis invasion out of ovaries	0.187	1.210 (0.790 to 1.840)	0.386
	rAFS staging	0.455	1.580 (1.010 to 2.470)	0.047
Postoperative variables	Management with GnRH-a	0.486	1.630 (0.994 to 2.66)	0.052
	Number of deliveries after surgery	−0.371	0.690 (0.470 to 1.010)	0.059
	Number of abortions after surgery	0.0003	1.000 (0.566 to 1.770)	0.999

Abbreviations: CA125, Carbohydrate antigen 125; CA199, Carbohydrate antigen 199; CRP, C-reactive protein; rAFS, Revised American Fertility Society.

Establishment of Multivariate Cox Regression Model

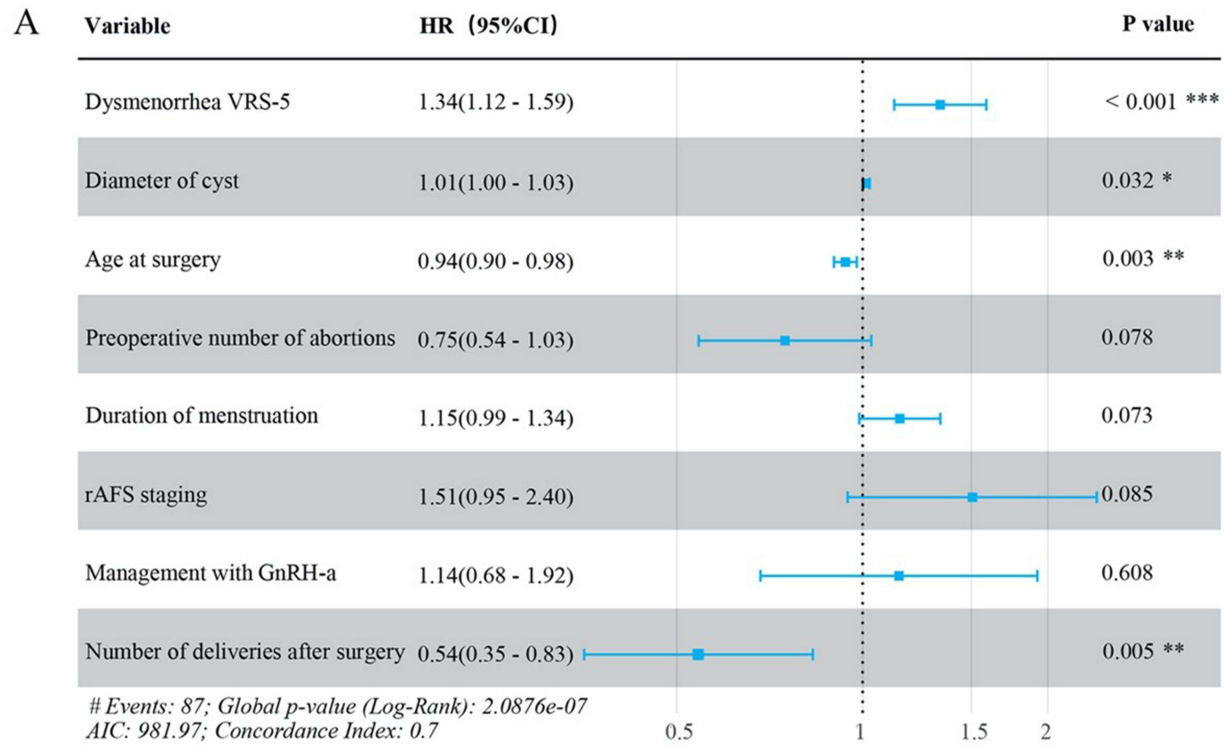
The variables with $p < 0.1$ in the univariate Cox analysis included dysmenorrhea VRS-5, diameter of ovarian cyst, age at the time of surgery, number of preoperative abortions, days of menstruation, intraoperative rAFS, postoperative GnRH-a treatment, and postoperative number of deliveries. These eight variables were subsequently included in multivariate Cox analysis for model construction. Among them, VRS-5 of dysmenorrhea, diameter of ovarian cyst, age at the time of surgery, and number of postoperative deliveries were identified as independent factors related to the recurrence of ovarian endometrioma (Figure 1A). Furthermore, the "rms" package was utilized to establish a nomogram for predicting 3-year and 5-year recurrence of postoperative ovarian endometrioma (Figure 1B).

Variables Were Screened in Random Survival Model

The model achieved stability when constructing 1000 survival trees after parameter tuning (Figure 2A). Using the VIMP method and the minimum depth method, 12 important variables were identified, including: VRS-5 of dysmenorrhea, age at surgery, maximum diameter of ovarian cyst, age of menarche, menstrual days, menstrual cycles, number of preoperative abortions, serum levels of triglycerides, fasting blood glucose, white blood cell counts, CRP, and CA199 (Figure 2B). The random survival forest model was established by incorporating these 12 variables. The model was constructed using the "mlr3" package in R, with hyperparameters fine-tuned to optimize performance. Model performance was assessed using the C-index. After parameter tuning, the best model parameters were selected as "mtry = 1", "nsplit = 50", and "nodesize = 1".

Validation and Comparison of Multivariate Cox Regression and Random Survival Forest Models

The Cox model exhibited a lower C-index compared to the RSF model in the training set (0.699 vs. 0.987). This suggested that RSF outperformed the Cox model in predicting survival outcome within the training data, indicating better



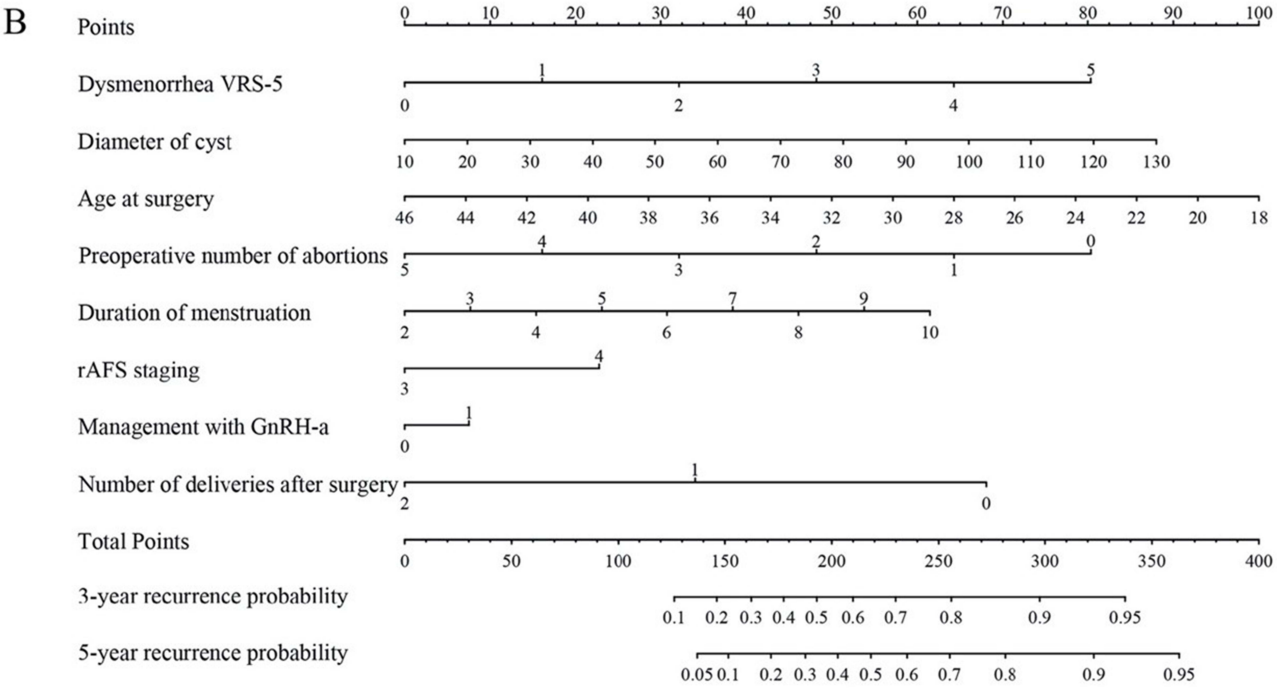


Figure 1 A nomogram of predicting recurrence of ovarian endometrioma. (**A**) Forest tree of hazard ratio of all variables. *** < 0.001; ** < 0.01; * < 0.05; (**B**) A nomogram of predicting 3-year and 5-year recurrence of ovarian endometrioma.

capture of underlying patterns in this dataset. Conversely, the Cox model had a slightly higher C-index compared to the RSF model in the validation set (0.771 vs. 0.763). Although the Cox model performed marginally better than the RSF model in predicting survival outcomes within the testing set, the difference was not as significant as observed in the

https://doi.org/10.2147/IJWH.S578547

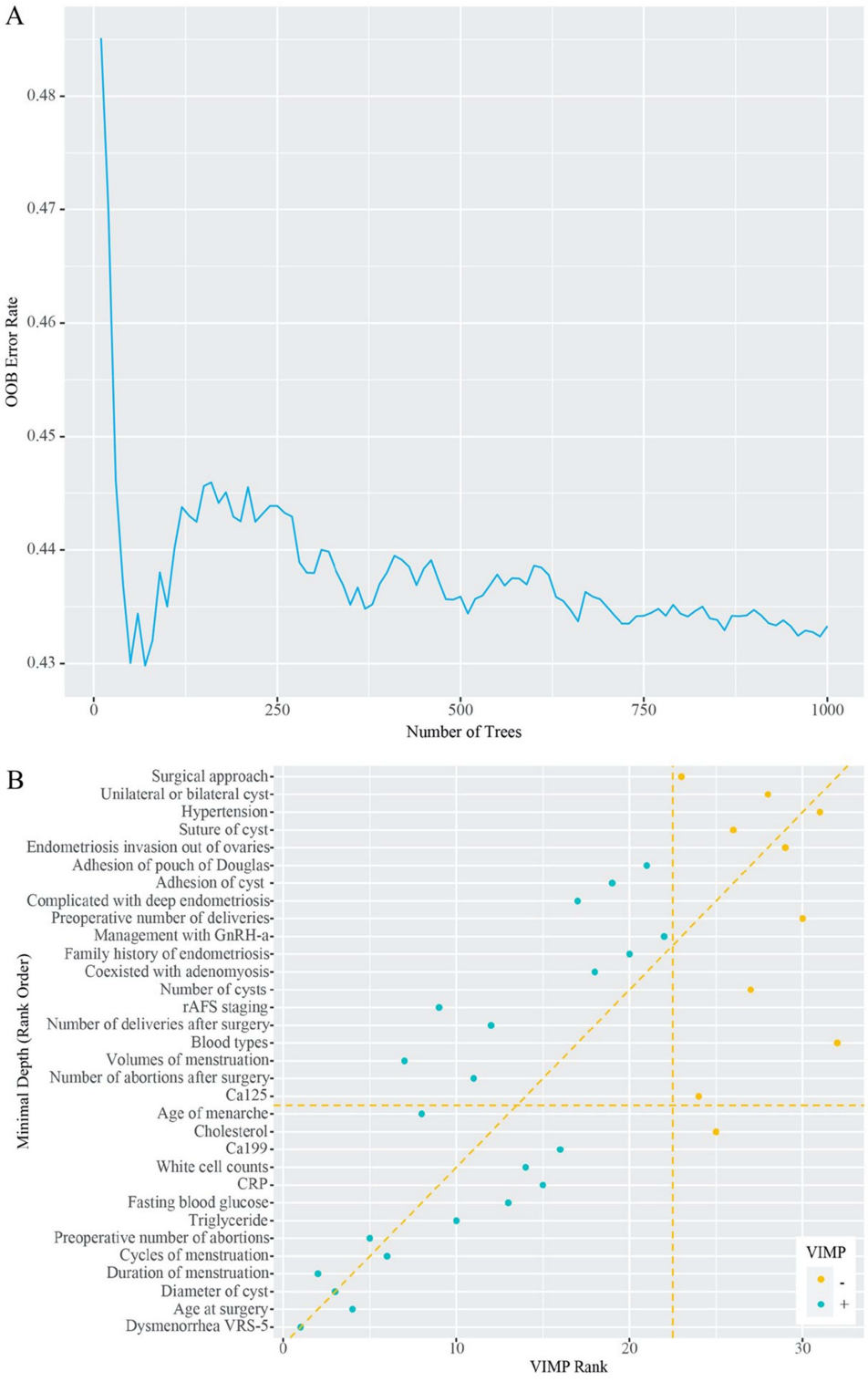


Figure 2 Random forest model for predicting recurrence of postoperative of ovarian endometrioma. (**A**) The stability of the model was achieved after constructing 1,000 survival trees through parameter debugging. (**B**) By employing the VIMP (Variable Importance in Projection) method and the minimum depth method, 12 significant variables were identified. In the graphical representation, the horizontal axis depicted the sorting of variables based on VIMP values (Blue dots: VIMP > 0, Red dots: VIMP < 0). The vertical axis was sorted by minimum depth, with variables closer to the lower-left corner considered more important.

training set. The 3-year AUC and 5-year AUC of the ROC curve for the Cox regression model were 0.723 (95% CI: 0.631–0.815) and 0.735 (95% CI: 0.662–0.808) respectively in the testing set (Figure 3A). Conversely, the 3-year AUC and 5-year AUC of the ROC curve for the RSF model were 0.997 (95% CI: 0.994–1.000) and 1.000 (95% CI: 0.999–1.000) respectively in the training set (Figure 3B). In the testing set, the 3-year AUC and 5-year AUC of the ROC curve for the Cox regression model were 0.784 (95% CI: 0.685–0.883) and 0.763 (95% CI: 0.653–0.873) respectively (Figure 3C), while for the RSF model, they were 0.792 (95% CI: 0.678–0.906) and 0.810 (95% CI: 0.712–0.908) respectively (Figure 3D). Furthermore, time-dependent ROC curves of the two models were generated for both the training and testing sets (Figure 3E–H). Calibration curves for 3-year and 5-year recurrence were validated and plotted using the Bootstrap method for both models in both the training and testing sets. (Figure 4).

In the decision curve analysis (DCA) of the prediction model, the area between each prediction model curve and the "None line" and "All line" indicated the clinical practicality of the model. The farther the distance between the model curve and the "None line" or "All line," the better the clinical value of the model (Figure 5).

The prediction model was validated through Kaplan-Meier curves. The risk scores of the Cox model and the RSF model were calculated in the training and testing set samples, and the optimal cutoff risk score value was determined using the "survminer" package in R to divide the population into high-risk and low-risk groups. Kaplan-Meier curves were then drawn, showing that the recurrence rate of high-risk patients was significantly higher than that of low-risk patients, with $p < 0.001$ (Figure 6A–D).

The classification accuracy of the RSF model and the Cox regression model were evaluated using two performance metrics: Net Reclassification Improvement (NRI) and Integrated Discrimination Improvement (IDI). The 3-year and 5-year NRIs were 0.728 (95% Confidence Interval (CI): 0.607 to 0.841) and 0.775 (95% CI: 0.649–0.853) respectively. The 3-year and 5-year IDIs were 0.599 (95% CI: 0.485–0.700) and 0.603 (95% CI: 0.502–0.698) respectively. These results suggest that the RSF model outperformed the traditional Cox regression model in terms of both NRI and IDI at both 3-year and 5-year intervals. (Figure 6E and F).

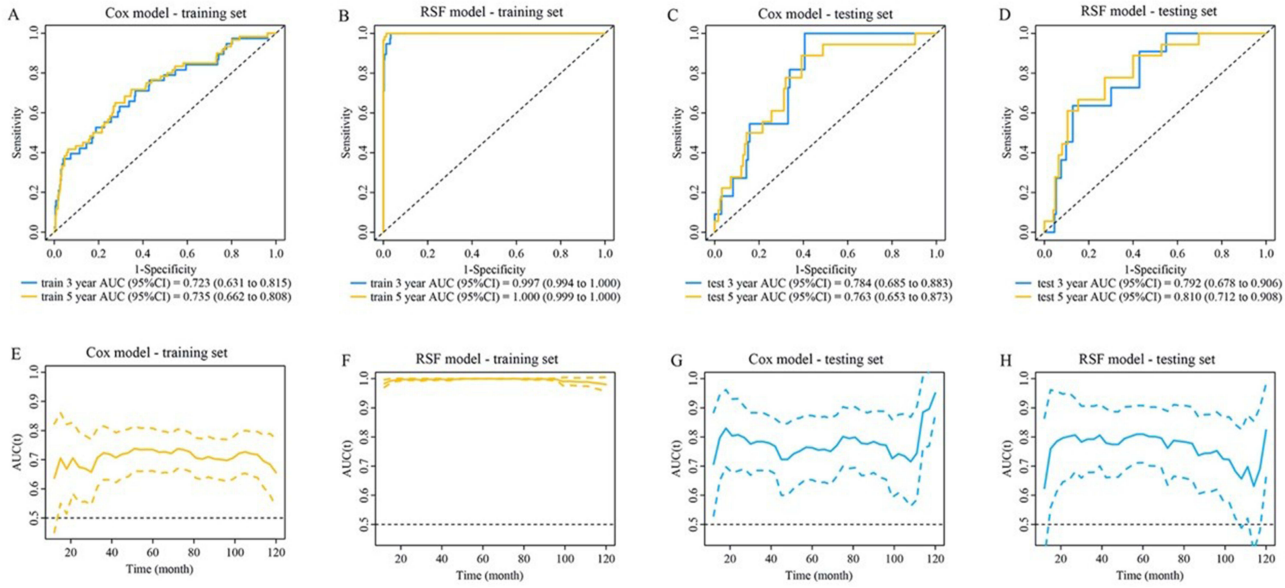


Figure 3 AUC and time-dependent ROC of models. (**A**) 3-year and 5-year AUC of Cox model in training set; (**B**) 3-year and 5-year AUC of RSF model in training set; (**C**) 3-year and 5-year AUC of Cox model in testing set; (**D**) 3-year and 5-year AUC of RSF model in testing set; (**E**) Time-dependent ROC curve of Cox model in training set (yellow); (**F**) Time-dependent ROC curve of RSF model in training set (yellow); (**G**) Time-dependent ROC curve of Cox model in testing set (blue); (**H**) Time-dependent ROC curve of RSF model in testing set (blue).

https://doi.org/10.2147/IJWH.S578547

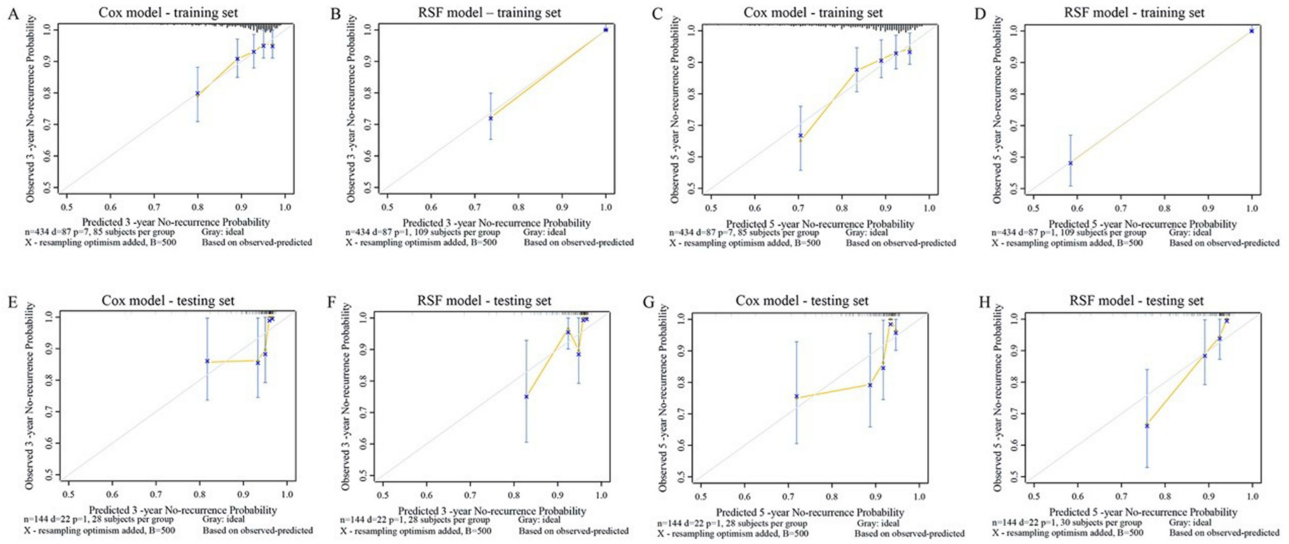


Figure 4 Calibration curves of models in the training and testing set. (**A**) 3-year calibration curve of Cox model in training set; (**B**) 3-year calibration curve of RSF model in training set; (**C**) 5-year calibration curve of Cox model in training set; (**D**) 5-year calibration curve of RSF model in training set; (**E**) 3-year calibration curve of Cox model in testing set; (**F**) 3-year calibration curve of RSF model in testing set. (**G**) 5-year calibration curve of Cox model in testing set; (**H**) 5-year calibration curve of RSF model in testing set.

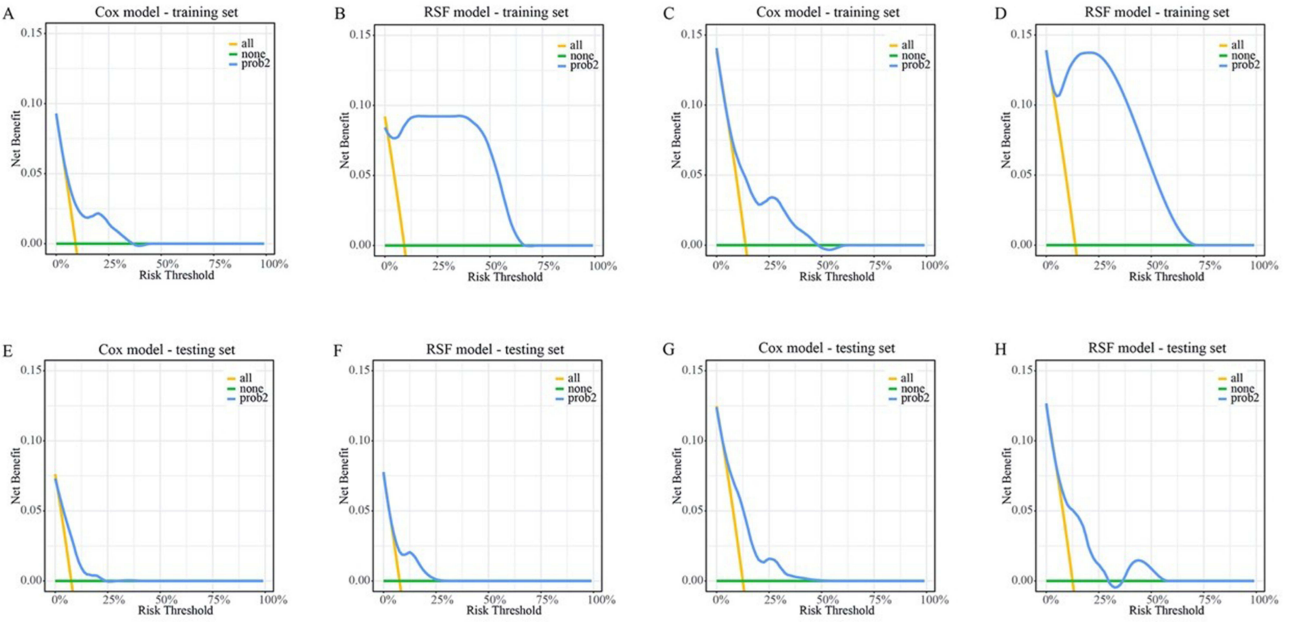


Figure 5 DCA curves of models in training and testing set. (**A**) 3-year DCA curve of Cox model in training set; (**B**) 3-year DCA curve of RSF model in training set; (**C**) 5-year DCA curve of Cox model in training set; (**D**) 5-year DCA curve of RSF model in training set; (**E**) 3-year DCA curve of Cox model in testing set; (**F**) 3-year DCA curve of FSF model in testing set; (**G**) 5-year DCA curve of Cox model in testing set; (**H**) 5-year DCA curve of RSF model in testing set.

Discussion

The postoperative recurrence rate of 18.86% among 578 patients was a significant finding in the present study, especially if it indeed represented a lower recurrence rate compared to previous studies. The wide range in recurrence rates, from 15% to 56%, underscores the variability in patient outcomes and possibly reflects differences in study designs, patient populations, surgical techniques, and postoperative management strategies.[5] Compared with previous studies, this study

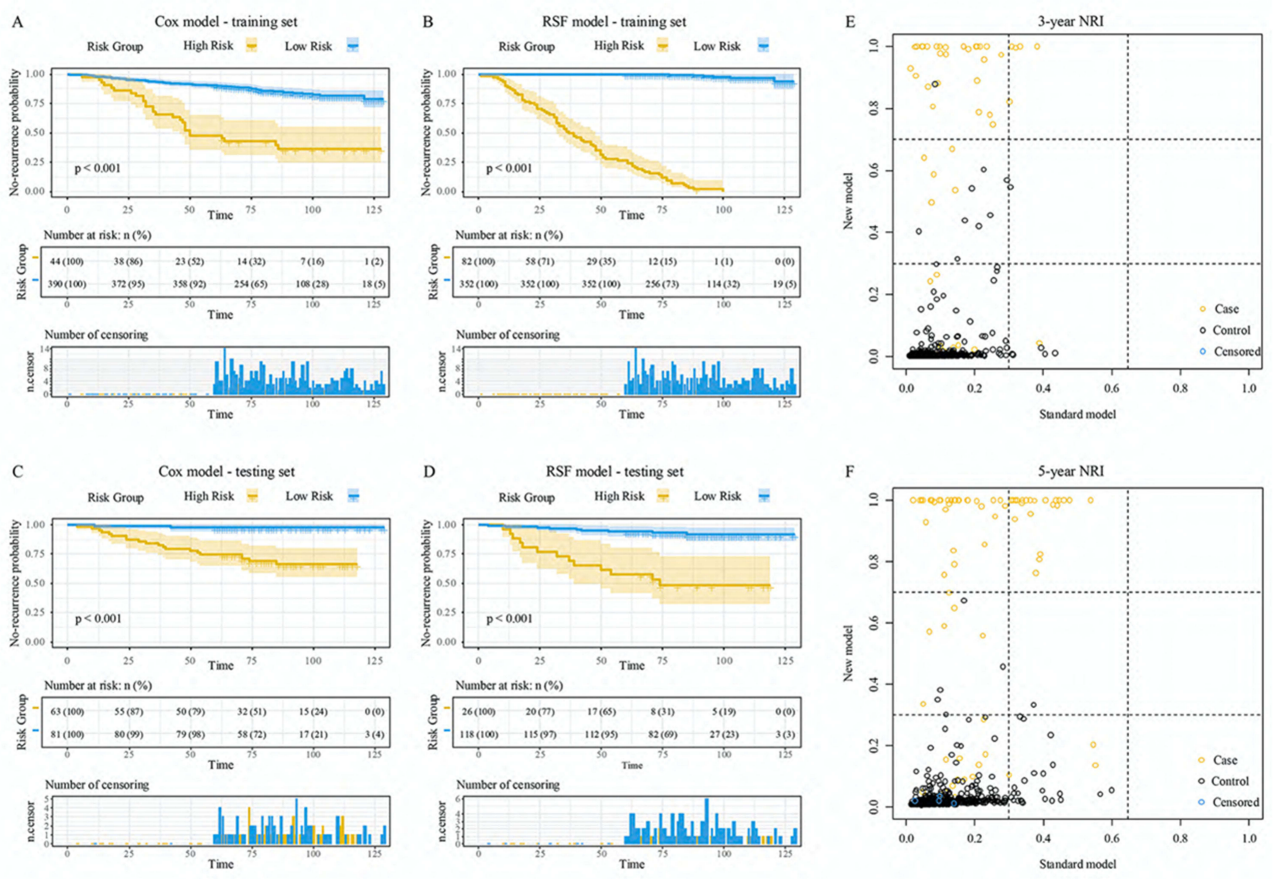


Figure 6 K-M curves and NRI of models. (**A**) K-M curve of Cox model in training set; (**B**) K-M curve of RSF model in training set; (**C**) K-M curve of Cox model in testing set; (**D**) K-M curve of RSF model in testing set; Compared to traditional Cox regression models, the RSF model was a 3-year NRI of 0.728 (0.607–0.841) (**E**) and a 5-year NRI of 0.775 (0.649–0.853) (**F**).

considered the patient's recurrence status and recurrence time. The findings that the RSF model outperformed the traditional Cox regression model in terms of partitioning and calibration in the present study suggested that RSF had good application value in predicting postoperative recurrence of ovarian endometrioma. The common important variables screened based on the two algorithms included VRS-grade of dysmenorrhea, age at the time of surgery, diameter of ovarian cyst, number of preoperative abortions, and days of menstruation. Among them, the VRS-5 of dysmenorrhea, age at the time of surgery, and ovarian cyst diameter were independent factors in Cox multivariate analysis. These independent factors were consistent with previous study.[6] Research has found that dysmenorrhea is the most prominent symptom of endometriosis.[7,8] Approximately 40–50% of patients experienced a recurrence of dysmenorrhea within five years after surgery.[9] As the age at surgery increased, the likelihood of postoperative recurrence decreased. Ovarian endometrioma ablation may reduce the recurrence rate of endometriomas by decreasing follicular reserve and hormone secretion, particularly in patients over 35 years old.[10] Larger ovarian endometriomas may indicate a more severe form of the condition or a greater capacity for the lesions to invade and grow within the ovary. The adverse effects of resection of ovarian endometrioma on ovarian function have been widely reported.[11–13] The diameter of ovarian endometrioma larger than 4 cm led to disruption of ovarian tissue, distortion of ovarian anatomy and reduction of the dominant follicles and oocytes of the ovary.[14] Therefore, for women with ovarian endometriomas who desire fertility preservation, careful consideration of the size of the cysts is important in treatment strategies.

In some cases, a conservative approach with medical management or fertility-sparing surgery may be considered to minimize the impact on ovarian reserve while still addressing symptoms and preserving fertility. Several important variables were identified by RSF, including age of menarche, menstrual cycle length, serum levels of triglycerides,

https://doi.org/10.2147/IJWH.S578547

fasting blood glucose, white blood cell count, CRP, and CA199. Among these variables, earlier age of menarche was found to be potentially associated with an increased risk of endometriosis.[15] This may be due to increased exposure to the hormonal environment and a higher likelihood of menstrual blood reflux. Additionally, studies have found that shorter menstrual cycles are associated with higher rates of endometriosis.[16,17] Several studies have indeed found a negative correlation between body mass index (BMI) and the occurrence of endometriosis.[18–20] This study did not find a direct relationship between BMI and endometriosis, but the RSF model identified triglycerides and fasting blood glucose as potential indicators that may be related to BMI. This suggested there could be indirect connections or underlying factors linking these variables. Genetic changes associated with transformation can lead to estrogen production and conversion.[21] Overall, the periodic bleeding from ectopic endometrial tissue in patients with endometriosis leads to oxidative stress and immune activation. As a result of the oxidative stress and immune response triggered by the presence of ectopic endometrial tissue, changes in the levels of white blood cells (which are involved in the immune response) and CRP, a marker of inflammation in the body, could occur.[22,23] Currently, there were no universally accepted laboratory biomarkers for endometriosis.[24,25] Pathological examination after surgery remained a crucial component of the diagnostic process. The most representative markers currently used in endometriosis included CA125 and CA199, which had similar specificities and responses to disease severity.[26,27] One study found a significant relationship between high expression of CA 199 and postoperative recurrence rate in endometriosis patients, suggesting its potential utility in predicting disease progression. GnRH-a was initially hypothesized to be an important factor affecting recurrence; however, it did not emerge as a significant variable after conducting two different analysis methods in this study. Yang et al had similar results through a cross-sectional study of 874 patients with ovarian endometrioma.[28] Research on hormone therapy for endometriosis found varying degrees of effectiveness, with approximately 66.67% of patients experiencing significant relief from symptoms and others experiencing little effect or relapse after treatment cessation.[29]

Indeed, this study had several notable strengths. Large sample size (578 cases) and long follow-up period (Mean: 90 months, ranging from 60 months to 129 months) were two significant strengths. By integrating the dimension of recurrence time into the predictive model of ovarian endometrioma recurrence using stochastic survival machine learning, the study introduced a novel approach that holds promise for enhanced accuracy, comprehensiveness, and clinical utility. This study is subject to several limitations that warrant consideration. Firstly, the retrospective nature of the study introduces the possibility of bias, potentially impacting the accuracy and reliability of the results. Furthermore, the study's exclusive focus on patients who underwent surgical treatment may result in a skewed representation of the overall endometriosis population, as it excludes individuals who opted for non-surgical interventions or did not seek medical care.

Conclusions

The RSF model exhibited a better fitting effect in the training set compared to the Cox regression model. However, there were no discernible differences in the predictive performance of the two models when assessing the recurrence of ovarian endometrioma in the testing set.

Data Sharing Statement

Additional data are available from corresponding author upon request.

Ethical Approval and Consent to Participate

This study was approved by the institutional review board of Suzhou Municipal Hospital (KL901451). The requirement for informed consent was waived by the institutional review board because this is retrospective study. The present work was conducted in accordance with the Helsinki Declaration. Patient data was confidential and anonymized.

Author Contributions

AW and SH contributed equally as corresponding authors. YZ, ML and ZZ contributed to this article equally. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing

the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors have no competing interests to declare in this work.

References

1. Saunders PTK, Horne AW. Endometriosis: etiology, pathobiology, and therapeutic prospects. *Cell*. 2021;184(11):2807–2824. doi:10.1016/j.cell.2021.04.041
2. Baraki D, Richards EG, Falcone T. Treatment of endometriomas: surgical approaches and the impact on ovarian reserve, recurrence, and spontaneous pregnancy. *Best Pract Res Clin Obstet Gynaecol*. 2024;92:102449. doi:10.1016/j.bpobgyn.2023.102449
3. Maul LV, Morrision JE, Schollmeyer T, Alkatout I, Mettler L. Surgical therapy of ovarian endometrioma: recurrence and pregnancy rates. *JSLS*. 2014;18(3):e2014–00223. doi:10.4293/JSLS.2014.00223
4. Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res JMLR*. 2019;20:177. doi:10.1080/01621459.1963.10500830
5. Ceccaroni M, Bounous VE, Clarizia R, Mautone D, Mabrouk M. Recurrent endometriosis: a battle against an unknown enemy. *Eur J Contracept Reprod Health Care off J Eur Soc Contracept*. 2019;24(6):464–474. doi:10.1080/13625187.2019.1662391
6. Ngernprom P, Klangsin S, Suwanrath C, Peeyananjarassri K. Risk factors for recurrent endometriosis after conservative surgery in a quaternary care center in southern Thailand. *PLoS One*. 2023;18(8):e0289832. doi:10.1371/journal.pone.0289832
7. Lee DY, Kim HJ, Yoon BK, Choi D. Factors associated with the laterality of recurrent endometriomas after conservative surgery. *Gynecol Endocrinol off J Int Soc Gynecol Endocrinol*. 2013;29(11):978–981. doi:10.3109/09513590.2013.824959
8. Sampson JA. Metastatic or embolic endometriosis, due to the menstrual dissemination of endometrial tissue into the venous circulation. *Am J Pathol*. 1927;3(2):93–110.43.
9. Guo SW. Recurrence of endometriosis and its control. *Hum Reprod Update*. 2009;15(4):441–461. doi:10.1093/humupd/dmp007
10. Shi J, Leng J, Cui Q, Lang J. Follicle loss after laparoscopic treatment of ovarian endometriotic cysts. *Int J Gynaecol Obstet off Organ Int Fed Gynaecol Obstet*. 2011;115(3):277–281. doi:10.1016/j.ijgo.2011.07.026
11. Nankali A, Kazeminia M, Jamshidi PK, et al. The effect of unilateral and bilateral laparoscopic surgery for endometriosis on Anti-Mullerian Hormone (AMH) level after 3 and 6 months: a systematic review and meta-analysis. *Health Qual Life Outcomes*. 2020;18(1):314. doi:10.1186/s12955-020-01561-3
12. Younis JS, Shapso N, Ben-Sira Y, Nelson SM, Izhaki I. Endometrioma surgery-a systematic review and meta-analysis of the effect on antral follicle count and anti-Müllerian hormone. *Am J Obstet Gynecol*. 2022;226(1):33–51.e7. doi:10.1016/j.ajog.2021.06.102
13. Kitajima M, Khan KN, Harada A, et al. Association between ovarian endometrioma and ovarian reserve. *Front Biosci Elite Ed*. 2018;10(1):92–102. doi:10.2741/e810
14. Somigliana E, Li Piani L, Paffoni A, et al. Endometriosis and IVF treatment outcomes: unpacking the process. *Reprod Biol Endocrinol RBE*. 2023;21:107. doi:10.1186/s12958-023-01157-8
15. Nnoaham KE, Webster P, Kumbang J, Kennedy SH, Zondervan KT. Is early age at menarche a risk factor for endometriosis? A systematic review and meta-analysis of case-control studies. *Fertil Steril*. 2012;98(3):702–712.e6. doi:10.1016/j.fertnstert.2012.05.035
16. Missmer SA, Hankinson SE, Spiegelman D, et al. Reproductive history and endometriosis among premenopausal women. *Obstet Gynecol*. 2004;104(5 Pt 1):965–974. doi:10.1097/01.AOG.0000142714.54857.f8
17. Shafrir AL, Farland LV, Shah DK, et al. Risk for and consequences of endometriosis: a critical epidemiologic review. *Best Pract Res Clin Obstet Gynaecol*. 2018;51:1–15. doi:10.1016/j.bpobgyn.2018.06.001
18. Parasar P, Ozcan P, Terry KL. Endometriosis: epidemiology, diagnosis and clinical management. *Curr Obstet Gynecol Rep*. 2017;6(1):34–41. doi:10.1007/s13669-017-0187-1
19. Shahbazi S, Shahrabi-Farahani M. Evaluation of the correlation between body mass index and endometriosis among Iranian fertile women. *Gynecol Endocrinol off J Int Soc Gynecol Endocrinol*. 2016;32(2):157–160. doi:10.3109/09513590.2015.1101439
20. Farland LV, Missmer SA, Bijon A, et al. Associations among body size across the life course, adult height and endometriosis. *Hum Reprod Oxf Eng*. 2017;32(8):1732–1742. doi:10.1093/humrep/dex207
21. Cardoso JV, Abrão MS, Berardo PT, et al. Role of cytochrome P450 2C19 polymorphisms and body mass index in endometriosis: a case-control study. *Eur J Obstet Gynecol Reprod Biol*. 2017;219:119–123. doi:10.1016/j.ejogrb.2017.10.027
22. Tarokh M, Ghaffari Novin M, Poordast T, et al. Serum and peritoneal fluid cytokine profiles in infertile women with endometriosis. *Iran J Immunol IJI*. 2019;16(2):151–162. doi:10.22034/IJI.2019.80258
23. Samimi M, Pourhanifeh MH, Mehdizadehkashi A, Eftekhar T, Asemi Z. The role of inflammation, oxidative stress, angiogenesis, and apoptosis in the pathophysiology of endometriosis: basic science and new insights based on gene expression. *J Cell Physiol*. 2019;234(11):19384–19392. doi:10.1002/jcp.28666
24. Smolarz B, Szyłło K, Romanowicz H. Endometriosis: epidemiology, classification, pathogenesis, treatment and genetics (Review of literature). *Int J Mol Sci*. 2021;22(19):10554. doi:10.3390/ijms221910554
25. Nisenblat V, Bossuyt PMM, Shaikh R, et al. Blood biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane Database Syst Rev*. 2016;2016(5):CD012179. doi:10.1002/14651858.CD012179

26. Rokhgireh S, Mehdizadeh Kashi A, Chaichian S, et al. The diagnostic accuracy of combined Enolase/Cr, CA125, and CA19-9 in the detection of endometriosis. *BioMed Res Int*. 2020;2020:5208279. doi:10.1155/2020/5208279
27. Harada T, Kubota T, Aso T. Usefulness of CA19-9 versus CA125 for the diagnosis of endometriosis. *Fertil Steril*. 2002;78(4):733–739. doi:10.1016/s0015-0282(02)03328-9
28. Yang X, Bao M, Hang T, et al. Status and related factors of postoperative recurrence of ovarian endometriosis: a cross-sectional study of 874 cases. *Arch Gynecol Obstet*. 2023;307(5):1495–1501. doi:10.1007/s00404-023-06932-x
29. Perrone U, Evangelisti G, Laganà AS, et al. A review of phase II and III drugs for the treatment and management of endometriosis. *Expert Opin Emerg Drugs*. 2023;28(4):333–351. doi:10.1080/14728214.2023.2296080