Am J Cancer Res 2025;15(8):3376-3394
www.ajcr.us /ISSN:2156-6976/ajcr0166182

Original Article

A high-risk prediction model for endometrial cancer: exploring the synergistic interaction between polycystic ovary syndrome and metabolic syndrome

Qian Xu[1], Xue Wu[2], Yabin Guo[2]

[1]Department of Gynecology, Changde Hospital, Xiangya School of Medicine, Central South University (The First People's Hospital of Changde City), Changde 415000, Hunan, China; [2]Center of Reproductive Medicine, Changde Hospital, Xiangya School of Medicine, Central South University (The First People's Hospital of Changde City), Changde 415000, Hunan, China

Received May 20, 2025; Accepted August 1, 2025; Epub August 15, 2025; Published August 30, 2025

Abstract: Objective: To investigate the synergistic interaction between polycystic ovary syndrome (PCOS) and metabolic syndrome (MetS) in relation to the risk of endometrial cancer (EC). Additionally, we aimed to develop a clinically applicable, high-risk early-warning model that incorporates these interactive factors, enhancing the precision and clinical utility of EC screening. Methods: We conducted a retrospective case-control study involving 445 newly diagnosed EC patients and 299 healthy female controls from the First People's Hospital of Changde City, between January 2018 and January 2025. Multivariate logistic regression was used to assess the independent and combined effects of PCOS and MetS on EC risk. A nomogram-based predictive model was developed and validated rigorously using training, internal validation, and external validation cohorts. The model's performance was evaluated based on discrimination (area under the curve [AUC]), calibration (Hosmer-Lemeshow test), and clinical utility (decision curve analysis). The diagnostic performance of our comprehensive model was compared to traditional tumor markers (cancer antigen 125/199, human epididymis protein 4). Results: LASSO regression identified 14 clinically significant predictors. Logistic regression revealed that HE4 levels, endometrial thickness, and fasting blood glucose were independent risk factors for EC, while high-density lipoprotein was an independent protective factor. The nomogram based on these variables demonstrated excellent discrimination, with AUCs of 0.984 in the training set, 0.987 in the internal validation set, and 0.964 in the external validation set. The integrated risk model significantly outperformed individual markers in diagnostic accuracy across all datasets ($P<0.001$). Conclusion: Our PCOS-MetS interaction-based EC risk prediction model showed robust and consistent performance across multiple validation cohorts. This tool significantly improves early detection accuracy and holds substantial clinical promise, laying the foundation for personalized EC risk management strategies.

Keywords: Endometrial carcinoma, polycystic ovary syndrome, metabolic syndrome, risk prediction model, nomograms

Introduction

Endometrial cancer (EC) is one of the most prevalent malignancies of the female reproductive system, with its global incidence steadily increasing, particularly in industrialized nations and developed regions [1]. According to the Global Cancer Statistics 2022, approximately 420,000 new EC cases were diagnosed, accounting for 4.5% of all female cancers, along with 97,000 related deaths, highlighting the significant burden on women's health and quality of life [2]. Evolving lifestyle patterns and the global obesity epidemic have contributed to the alarming trend of earlier disease onset, suggesting a strong association between EC risk factors and modern living conditions [3]. The pathogenesis of EC is multifactorial, involving hormonal imbalances, metabolic dysregulation, and genetic predisposition. Of particular clinical relevance, polycystic ovary syndrome (PCOS) [4] and metabolic syndrome (MetS) [5], two common endocrine-metabolic disorders - have been increasingly recognized as major risk factors for endometrial carcinogenesis in recent epidemiological and pathophysiological studies.

https://doi.org/10.62347/ZZPA6435

PCOS, a common endocrine disorder affecting 5-10% of reproductive-aged women, is clinically characterized by hyperandrogenism, ovulatory dysfunction, and polycystic ovarian morphology [6]. A hallmark of PCOS is chronic anovulation and a hyperestrogenic state, with prolonged estrogen exposure being mechanistically linked to endometrial carcinogenesis [7]. Ignatov et al. [8] demonstrated that PCOS-driven hyperandrogenism and unopposed estrogen stimulation substantially increase EC risk. Epidemiological evidence indicates that women with PCOS have a threefold higher risk of EC compared to non-PCOS women, with an especially pronounced risk in premenopausal populations [9]. Compounding this risk, PCOS frequently coexists with metabolic disturbances, including insulin resistance, obesity, and dyslipidemia - features that overlap with MetS, further promoting endometrial carcinogenesis [10].

MetS is a cluster of interrelated metabolic disturbances, including obesity, glucose intolerance, hypertension, and atherogenic dyslipidemia, with insulin resistance and chronic low-grade inflammation serving as its pathophysiological foundation [11]. The global prevalence of MetS has risen dramatically, paralleling the obesity epidemic [12]. Substantial evidence links MetS components to EC pathogenesis, particularly through mechanisms related to insulin resistance [13]. For example, a prospective cohort study in Chinese women reported a 1.8-fold higher EC incidence in MetS patients, with risk increasing in a dose-dependent manner as the number of metabolic abnormalities rose [13]. The underlying oncogenic mechanisms include hyperactivation of the insulin-like growth factor 1 (IGF-1) pathway and sustained pro-inflammatory cytokine release, both of which drive dysregulated endometrial proliferation and malignant transformation [14].

Notably, there is considerable overlap between PCOS and MetS in patient populations, suggesting potential interactions between these conditions that may collectively contribute to EC development [15]. For instance, PCOS-related hyperandrogenism and elevated estrogen levels may interact with MetS-associated insulin resistance and chronic inflammation, creating a feedback loop that further exacerbates EC risk [16]. Studies have shown that obesity and dyslipidemia in PCOS patients - traits that overlap with MetS - significantly increase the risk of endometrial hyperplasia and carcinogenesis [17]. However, research on the precise mechanistic interaction between PCOS and MetS, especially large-scale, multi-dataset validated analyses, remains limited. Therefore, elucidating the interaction mechanisms between PCOS and MetS and developing a robust, risk-stratified early-warning model for EC based on these insights would have substantial theoretical and clinical importance.

This study aims to develop and validate a high-risk early-warning model for EC by integrating the synergistic interaction between PCOS and MetS, thereby improving the accuracy and clinical utility of early screening. Using logistic regression analysis, we will assess the independent effect of the PCOS-MetS interaction on EC risk and construct a nomogram to quantify individual risk. The model's performance will be rigorously evaluated through multi-stage validation, including training, internal, and external validation datasets, to assess its discriminatory ability, calibration, and clinical net benefit. Additionally, we will compare the diagnostic value of this integrated model with that of individual tumor biomarkers to determine its superiority in risk prediction. The findings from this study will provide a theoretical basis for future individualized risk management strategies in patients with PCOS and MetS, addressing current research gaps and contributing to the advancement of precision medicine in EC prevention and early detection.

Methods and materials

Sample size calculation

The required sample size was determined using the pmsampsize package in R (type="b"), based on parameters derived from a prior study by Kuai et al. [18]. The Cox-Snell R-squared value was approximated using the c-statistic, with an area under the curve (AUC) of 0.863. Other inputs included a predicted event rate of 7, an outcome incidence rate of 61.61%, and an estimated shrinkage factor of 90%. The initial calculation indicated a minimum sample size of 364 participants. To account for potential data loss (10% attrition), the final sample size was adjusted to 404.

Study population

This retrospective study enrolled 445 patients with a primary diagnosis of EC, treated at the First People's Hospital of Changde City between January 2018 and January 2025. For comparison, 299 age-matched healthy women who underwent routine health examinations during the same period were included as controls. The study protocol was approved by the Institutional Review Board of The First People's Hospital of Changde city (**Figure 1**).

Sample grouping

The sample collection period spanned from January 2018 to January 2023. A total of 595 samples were collected, with 416 assigned to the training set and 179 to the validation set, in a 7:3 ratio. The training set was used to build and train the model, while the validation set assessed the model's accuracy and generalizability. Additionally, the external validation set was collected from February 2023 to January 2025, comprising 149 samples, to evaluate the model's performance and reliability across different datasets.

Inclusion and exclusion criteria

Inclusion criteria: Female sex; age ≥18 years; histopathologically confirmed EC [19]; PCOS diagnosed per the Rotterdam criteria [20]; and availability of complete clinical records.

Exclusion criteria: Prior history of malignancy (non-EC); recent chemotherapy/radiotherapy; severe hepatic or renal impairment; current pregnancy or lactation; or recent use of hormonal therapy.

Indicator collection

All data were extracted from patients' electronic medical records, including inpatient records, surgical documentation, imaging reports, and laboratory test results. Outpatient records, such as consultation notes and examination records, were also incorporated to ensure data reliability and completeness.

Clinical data: The following parameters were collected: age (≥55 years), body mass index (BMI, categorized as <23, 23-25, or >25), age at menarche (≥14 years), smoking history, diabetes history, hypertension history, abnormal uterine bleeding (AUB), gravidity (≥2 pregnancies), parity (≥1 delivery), menstrual cycle regularity (irregular or regular), polycystic features (assessed based on the Rotterdam criteria), and endometrial thickness (ET, measured via imaging).

Laboratory data: Fasting blood glucose (FBG), alanine aminotransferase (ALT), aspartate aminotransferase (AST), total protein (TP), albumin (ALB), globulin (GLO), total bilirubin (TBIL), uric acid (UA), free triiodothyronine (FT3), free thyroxine (FT4), thyroid-stimulating hormone (TSH), cancer antigen 125/199 (CA125/199), and human epididymis protein 4 (HE-4).

Lipid profile indicators: Triglycerides (TG), total cholesterol (TC), high-/low-density lipoprotein (HDL/LDL), apolipoprotein A1 (ApoA1), apolipoprotein B (ApoB), and lipoprotein(a) (Lp(a)).

Outcome measures

Primary outcomes: (1) Development and validation of a nomogram predictive model for EC. (2) Interaction analysis to evaluate the synergistic effects between key variables and EC risk.

Secondary outcomes: (1) Comparative analysis of baseline characteristics (e.g., age, BMI) across datasets. (2) Assessment of variable differences between the EC and control groups within the training set. (3) Feature selection using LASSO regression to identify significant predictors. (4) Logistic regression to determine the risk effects of variables on EC. (5) Binary classification via receiver operating characteristic (ROC) curve analysis to establish optimal cut-off values. (6) Comparative evaluation of tumor markers to demonstrate the superior predictive performance of the integrated model over individual markers.

Statistical analysis

All statistical analyses were performed using SPSS 27.0 and R 4.3.3, with statistical significance set at two-tailed $P<0.05$. Frequency distributions (n, %) were used to summarize categorical variables, and Pearson's chi-square tests (χ^2 tests) were applied for comparisons.

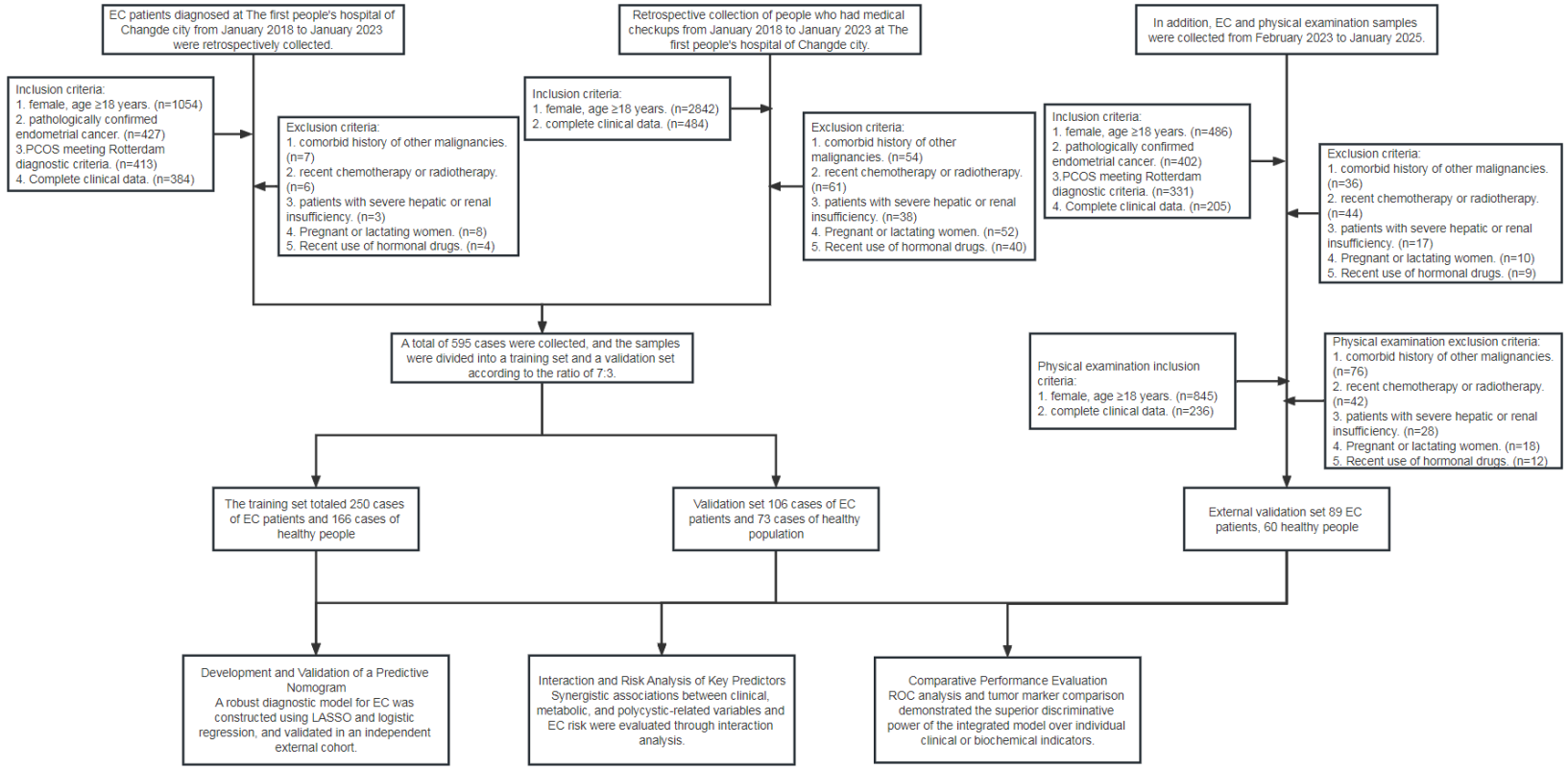

Figure 1. Sample collection flow chart. Note: EC, Endometrial Cancer; PCOS, Polycystic Ovary Syndrome.

Table 1. Baseline characteristics of the training, validation, and external validation sets

Variable	Training set (n=416)	Validation set (n=179)	External validation set (n=149)	Statistic	P
Age					
≥55	182 (43.75%)	93 (51.96%)	69 (46.31%)	3.390	0.184
<55	234 (56.25%)	86 (48.04%)	80 (53.69%)		
BMI					
<23	117 (28.12%)	60 (33.52%)	39 (26.17%)	3.157	0.532
23-25	142 (34.13%)	59 (32.96%)	57 (38.26%)		
>25	157 (37.74%)	60 (33.52%)	53 (35.57%)		
Age at menarche					
≥14	279 (67.07%)	124 (69.27%)	95 (63.76%)	1.125	0.570
<14	137 (32.93%)	55 (30.73%)	54 (36.24%)		
Smoking history					
Yes	105 (25.24%)	52 (29.05%)	38 (25.50%)	0.987	0.610
No	311 (74.76%)	127 (70.95%)	111 (74.50%)		
Diabetes history					
Yes	41 (9.86%)	19 (10.61%)	13 (8.72%)	0.330	0.848
No	375 (90.14%)	160 (89.39%)	136 (91.28%)		
Hypertension history					
Yes	107 (25.72%)	33 (18.44%)	28 (18.79%)	5.330	0.070
No	309 (74.28%)	146 (81.56%)	121 (81.21%)		
AUB					
Yes	240 (57.69%)	110 (61.45%)	93 (62.42%)	1.373	0.503
No	176 (42.31%)	69 (38.55%)	56 (37.58%)		
Gravidity					
≥2	300 (72.12%)	129 (72.07%)	109 (73.15%)	0.066	0.967
<2	116 (27.88%)	50 (27.93%)	40 (26.85%)		
Parity					
≥1	331 (79.57%)	146 (81.56%)	112 (75.17%)	2.109	0.348
<1	85 (20.43%)	33 (18.44%)	37 (24.83%)		
Menstrual cycle regularity					
Irregular	321 (77.16%)	139 (77.65%)	121 (81.21%)	1.075	0.584
Regular	95 (22.84%)	40 (22.35%)	28 (18.79%)		
Polycystic features					
Yes	46 (11.06%)	15 (8.38%)	15 (10.07%)	0.983	0.612
No	370 (88.94%)	164 (91.62%)	134 (89.93%)		
ET (cm)	0.67 [0.44, 0.94]	0.66 [0.40, 0.92]	0.67 [0.43, 1.01]	0.503	0.778
FBG (mmol/L)	5.91±0.85	5.96±0.88	5.97±0.87	0.313	0.731
ALT (U/L)	17.09±5.13	17.84±5.33	17.42±5.34	1.307	0.271
AST (U/L)	20.80 [18.20, 23.20]	20.30 [17.75, 23.25]	21.20 [18.90, 23.10]	1.700	0.427
TP (g/L)	74.69±7.97	74.78±7.80	74.48±7.57	0.063	0.939
ALB (g/L)	43.69±4.99	44.22±4.79	43.75±5.10	0.747	0.474
GLO (g/L)	29.81±3.90	29.84±4.44	30.45±3.88	1.466	0.232
TBIL (μmol/L)	11.20 [7.00, 14.95]	11.60 [8.10, 14.60]	10.10 [6.60, 14.20]	2.055	0.358
TG (mmol/L)	1.40 [1.19, 1.63]	1.44 [1.23, 1.61]	1.45 [1.16, 1.65]	0.701	0.704
TC (mmol/L)	4.90 [4.60, 5.23]	4.96 [4.61, 5.27]	4.88 [4.57, 5.26]	1.277	0.528
HDL (mmol/L)	1.21 [0.97, 1.46]	1.19 [0.96, 1.44]	1.27 [1.01, 1.45]	2.653	0.265
LDL (mmol/L)	3.01±0.53	3.02±0.51	3.03±0.49	0.133	0.875
ApoA1 (mg/dL)	138.42±29.09	136.51±29.76	139.40±25.86	0.455	0.635
ApoB (mg/dL)	97.24±21.93	96.26±21.50	94.70±23.17	0.734	0.480
Lp(a) (mg/L)	133.00 [113.00, 154.00]	136.00 [117.00, 155.00]	133.00 [112.00, 150.00]	2.510	0.285

UA (μmol/L)	287.45 [247.40, 326.72]	283.10 [248.85, 328.75]	286.00 [255.30, 331.70]	0.020	0.990
FT3 (pmol/L)	2.65±0.73	2.64±0.73	2.61±0.74	0.143	0.867
FT4 (pmol/L)	15.94±3.92	15.76±3.80	15.72±3.82	0.247	0.781
TSH (mIU/L)	2.24±0.61	2.29±0.59	2.23±0.64	0.561	0.571
CA125 (U/mL)	20.09±5.55	20.68±5.24	19.72±5.43	1.339	0.263
CA199 (U/mL)	16.26±5.34	15.86±5.48	15.78±5.26	0.633	0.531
HE-4 (pmol/L)	55.00 [37.00, 75.00]	54.00 [38.00, 73.00]	56.00 [35.00, 75.00]	0.033	0.984

Note: BMI, body mass index; AUB, abnormal uterine bleeding; ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TP, total protein; ALB, albumin; GLO, globulin; TBIL, total bilirubin; TG, triglycerides; TC, total cholesterol; HDL/LDL, high-/low-density lipoprotein; ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; Lp(a), lipoprotein(a); UA, uric acid; FT3, free triiodothyronine; FT4, free thyroxine; TSH, thyroid-stimulating hormone; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4.

For continuous variables, normality was assessed using the Kolmogorov-Smirnov test. Normally distributed data were expressed as mean ± standard deviation ($\bar{x}$ ± SD) and analyzed using independent samples t-tests. Nonparametric data were reported as median and interquartile range (IQR; M [Q1, Q3]) and tested using the Mann-Whitney U method.

Variable selection was performed using LASSO regression (glmnet package in R 4.3.3) with 10-fold cross-validation, identifying lambda.1se=0.016601 as the optimal penalty parameter, retaining 14 predictive variables from the initial 17 candidates. The discriminative performance of 11 continuous variables was evaluated through ROC curve analysis (pROC package in R 4.3.3), with cut-off values determined and AUC, sensitivity, and specificity reported. Univariate and multivariate logistic regression models were used to assess associations between selected variables and EC risk, with results presented as odds ratios (ORs) and 95% confidence intervals (CIs).

Synergistic interaction was examined using generalized linear models (stats package in R 4.3.3), with estimates and *P*-values reported. A nomogram incorporating 12 clinically relevant variables was developed to visualize the predictive model. Model performance was evaluated through ROC analysis assessing discriminative ability, calibration curves with 1000 bootstrap resamples (rms package), Hosmer-Lemeshow goodness-of-fit (GOF) test, and decision curve analysis (DCA) to evaluate clinical utility (rms package). The DeLong test (pROC package) was used to compare AUC differences between tumor marker-only models and the comprehensive predictive model.

Results

Comparison of baseline characteristics of the training, validation, and external validation sets

The training, validation, and external validation sets exhibited strong comparability across various baseline characteristics. No statistically significant differences (all P>0.05) were observed among the three sets for age, BMI, age at menarche, smoking history, diabetes history, AUB, gravidity, parity, menstrual cycle regularity, polycystic features, ET, FBG, liver function indices (ALT, AST, TP, ALB, GLO, TBIL), lipid profiles (TG, TC, HDL, LDL, ApoA1, ApoB, Lp(a)), UA levels, thyroid function indices (FT3, FT4, TSH), or tumor markers (CA125, CA199, HE-4). Although the difference in hypertension history among the sets approached statistical significance (P=0.070), it did not meet the threshold for significance. A detailed comparison is provided in **Table 1**.

Comparison of baseline characteristics between EC and control groups in the training cohort

Significant differences were observed in several baseline characteristics between the EC and control groups. BMI was significantly higher in the EC group (P=0.023), indicating a greater proportion of elevated BMI in these patients. Additionally, the EC group demonstrated significantly higher prevalence or levels of the following factors compared to controls: AUB (P<0.001), parity (P=0.014), menstrual cycle regularity (P=0.030), polycystic features (P< 0.001), ET (P<0.001), FBG (P<0.001), ALT (P<0.001), AST (P<0.001), TG (P<0.001), TC (P=0.035), HDL (P<0.001), LDL (P<0.001),

Table 2. Baseline characteristics of EC patients and controls in the training cohort

Variable	Total	EC group (n=250)	Control group (n=166)	Statistic	P
Age					
≥55	182 (43.75%)	107 (42.80%)	75 (45.18%)	0.230	0.632
<55	234 (56.25%)	143 (57.20%)	91 (54.82%)		
BMI					
<23	117 (28.12%)	59 (23.60%)	58 (34.94%)	7.586	0.023
23-25	142 (34.13%)	86 (34.40%)	56 (33.73%)		
>25	157 (37.74%)	105 (42.00%)	52 (31.33%)		
Age at menarche					
≥14	279 (67.07%)	166 (66.40%)	113 (68.07%)	0.126	0.722
<14	137 (32.93%)	84 (33.60%)	53 (31.93%)		
Smoking history					
Yes	105 (25.24%)	60 (24.00%)	45 (27.11%)	0.511	0.475
No	311 (74.76%)	190 (76.00%)	121 (72.89%)		
Diabetes history					
Yes	41 (9.86%)	29 (11.60%)	12 (7.23%)	2.145	0.143
No	375 (90.14%)	221 (88.40%)	154 (92.77%)		
Hypertension history					
Yes	107 (25.72%)	62 (24.80%)	45 (27.11%)	0.278	0.598
No	309 (74.28%)	188 (75.20%)	121 (72.89%)		
AUB					
Yes	240 (57.69%)	182 (72.80%)	58 (34.94%)	58.585	<0.001
No	176 (42.31%)	68 (27.20%)	108 (65.06%)		
Gravidity					
≥2	300 (72.12%)	185 (74.00%)	115 (69.28%)	1.107	0.293
<2	116 (27.88%)	65 (26.00%)	51 (30.72%)		
Parity					
≥1	331 (79.57%)	189 (75.60%)	142 (85.54%)	6.065	0.014
<1	85 (20.43%)	61 (24.40%)	24 (14.46%)		
Menstrual cycle regularity					
Irregular	321 (77.16%)	202 (80.80%)	119 (71.69%)	4.702	0.030
Regular	95 (22.84%)	48 (19.20%)	47 (28.31%)		
Polycystic features					
Yes	46 (11.06%)	40 (16.00%)	6 (3.61%)	15.560	<0.001
No	370 (88.94%)	210 (84.00%)	160 (96.39%)		
ET (cm)	0.67 [0.44, 0.94]	0.84 [0.60, 1.14]	0.51 [0.37, 0.64]	10.118	<0.001
FBG (mmol/L)	5.91±0.85	6.08±0.76	5.66±0.93	-4.978	<0.001
ALT (U/L)	17.09±5.13	18.49±4.88	14.98±4.78	-7.249	<0.001
AST (U/L)	20.55±3.99	22.26±3.09	17.97±3.82	-12.625	<0.001
TP (g/L)	74.69±7.97	74.83±8.02	74.49±7.92	-0.421	0.674
ALB (g/L)	43.69±4.99	43.93±4.96	43.34±5.02	-1.170	0.243
GLO (g/L)	29.81±3.90	29.79±3.78	29.83±4.08	0.097	0.923
TBIL (μmol/L)	11.20 [7.00, 14.95]	11.30 [7.10, 14.80]	10.90 [6.60, 15.45]	0.366	0.714
TG (mmol/L)	1.40 [1.19, 1.63]	1.52 [1.32, 1.69]	1.25 [1.02, 1.43]	8.919	<0.001
TC (mmol/L)	4.90 [4.60, 5.23]	4.94 [4.62, 5.28]	4.87 [4.56, 5.12]	2.113	0.035
HDL (mmol/L)	1.23±0.37	1.16±0.31	1.34±0.43	4.783	<0.001
LDL (mmol/L)	3.01±0.53	3.11±0.51	2.85±0.52	-5.097	<0.001
ApoA1 (mg/dL)	138.00 [121.00, 154.25]	135.50 [116.25, 150.75]	144.00 [124.25, 162.00]	3.611	<0.001
ApoB (mg/dL)	97.24±21.93	100.64±21.90	92.11±21.01	-3.954	<0.001
Lp(a) (mg/L)	133.01±30.26	133.05±30.27	132.96±30.33	-0.030	0.976
UA (μmol/L)	289.61±61.10	289.06±64.35	290.43±56.03	0.225	0.822

FT3 (pmol/L)	2.65±0.73	2.64±0.74	2.68±0.73	0.554	0.580
FT4 (pmol/L)	15.94±3.92	15.83±3.90	16.12±3.95	0.745	0.457
TSH (mIU/L)	2.24±0.61	2.22±0.59	2.27±0.63	0.905	0.366
CA125 (U/mL)	20.09±5.55	21.79±5.23	17.53±5.02	-8.258	<0.001
CA199 (U/mL)	16.26±5.34	18.25±4.89	13.27±4.54	-10.468	<0.001
HE-4 (pmol/L)	56.38±24.49	69.64±20.72	36.40±13.87	-18.140	<0.001

Note: EC, endometrial cancer; BMI, body mass index; AUB, abnormal uterine bleeding; ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TP, total protein; ALB, albumin; GLO, globulin; TBIL, total bilirubin; TG, triglycerides; TC, total cholesterol; HDL/LDL, high-/low-density lipoprotein; ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; Lp(a), lipoprotein(a); UA, uric acid; FT3, free triiodothyronine; FT4, free thyroxine; TSH, thyroid-stimulating hormone; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4.

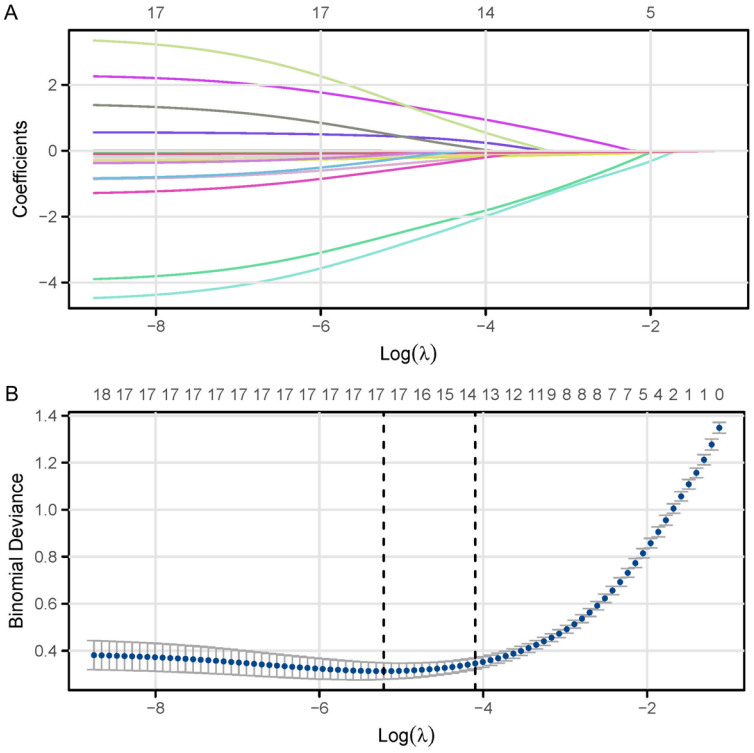

Figure 2. Feature selection process in LASSO regression. A. Variable trajectory plot showing feature selection. B. Coefficient profile plot demonstrating variable shrinkage.

ApoA1 (P<0.001), ApoB (P<0.001), CA125 (P< 0.001), CA199 (P<0.001), and HE-4 (P<0.001). In contrast, no statistically significant differences were observed between groups for age (P=0.632), age at menarche (P=0.722), smoking history (P=0.475), diabetes history (P= 0.143), hypertension history (P=0.598), gravidity (P=0.293), TP (P=0.674), ALB (P=0.243), GLO (P=0.923), TBIL (P=0.714), Lp(a) (P=0.976), UA (P=0.822), FT3 (P=0.580), FT4 (P=0.457), and TSH (P=0.366). For detailed numerical comparisons, please refer to **Table 2**.

Results of feature selection via LASSO regression

LASSO regression analysis was performed on 17 candidate variables, with the optimal penalty parameter (lambda.1se) set to 0.016601. This analysis identified 14 clinically relevant predictive variables: ET, FBG, ALT, AST, TG, HDL, LDL, ApoB, CA125, CA199, HE-4, AUB, menstrual cycle regularity, and polycystic features. These selected variables showed significant predictive value in the high-risk early-warning model for EC, as illustrated in **Figure 2A**, **2B**.

Conversion of continuous variables to binary categories for logistic regression analysis

To facilitate logistic regression analysis, optimal cutoff values for converting continuous variables into binary categories were determined using ROC curve analysis. The results demonstrated that ET, FBG, ALT, AST, TG, HDL, LDL, ApoB, CA125, CA199, and HE-4 exhibited significant discriminatory power, as evidenced by their respective AUC values. Among these biomarkers, HE-4 showed the highest predictive accuracy (AUC), followed by ET and AST. The optimal cutoff values, determined by the Youden index, were applied for binary classification, providing a strong foundation for logistic regression modeling (**Table 3** and **Figure 3**).

Table 3. ROC curve analysis of diagnostic performance for 11 continuous variables

Marker	AUC	95% CI	Specificity	Sensitivity	Youden index	Cut off
ET (mm)	0.793	0.750-0.836	92.17%	64.80%	56.97%	0.735
FBG (mmol/L)	0.638	0.582-0.694	49.40%	76.00%	25.40%	5.575
ALT (U/L)	0.7	0.649-0.751	72.89%	61.20%	34.09%	17.45
AST (U/L)	0.814	0.772-0.857	77.11%	73.20%	50.31%	20.55
TG (mmol/L)	0.758	0.711-0.805	53.61%	84.40%	38.01%	1.255
HDL (mmol/L)	0.619	0.562-0.676	42.77%	77.60%	20.37%	1.405
LDL (mmol/L)	0.643	0.589-0.697	59.64%	64.40%	24.04%	2.935
ApoB (mg/dL)	0.61	0.556-0.665	52.41%	66.40%	18.81%	91.500
CA125 (U/mL)	0.723	0.673-0.772	81.33%	56.00%	37.33%	21.090
CA199 (U/mL)	0.772	0.727-0.817	75.30%	67.20%	42.50%	16.170
HE-4 (pmol/L)	0.907	0.879-0.934	90.96%	75.60%	66.56%	55.500

Note: ROC, receiver operating characteristic; AUC, ara under the curve; CI, confidence interval; ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; HDL/LDL, high-/low-density lipoprotein; ApoB, apolipoprotein B; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4.

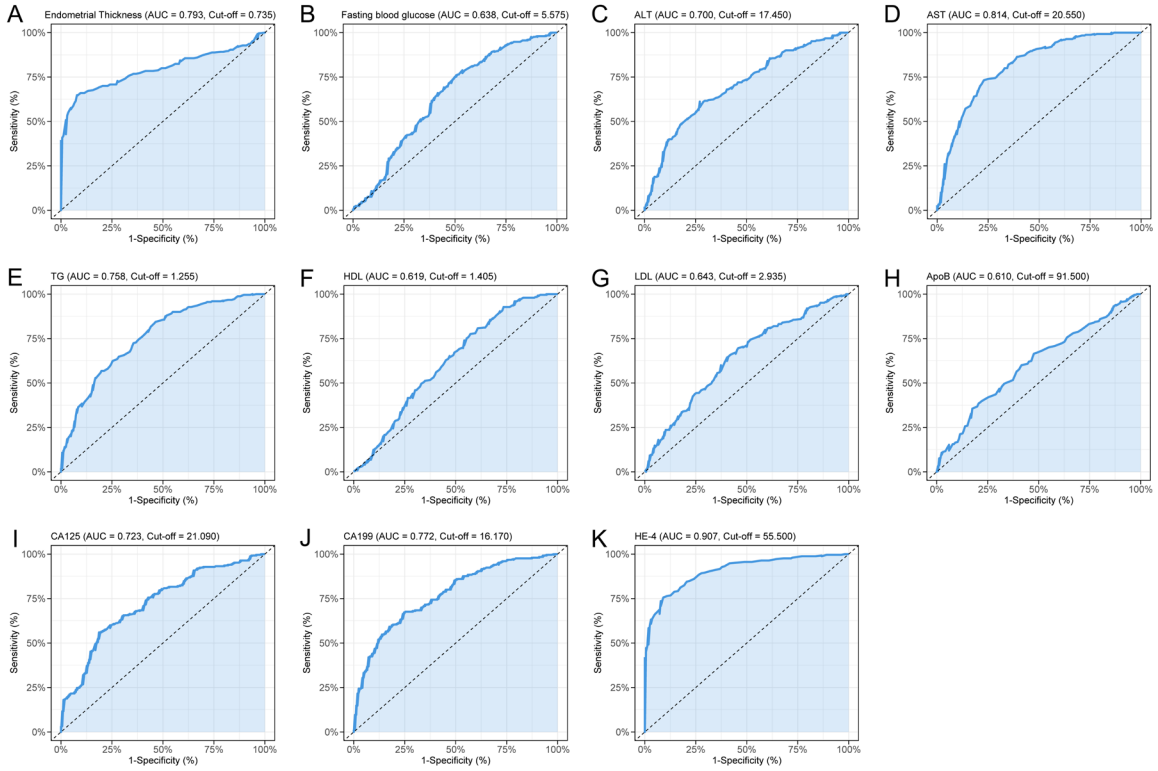

Figure 3. Determination of optimal binary classification cutoffs using ROC curve analysis. ROC curves illustrating the diagnostic performance of (A) ET, (B) FBG, (C) ALT, (D) AST, (E) TG, (F) HDL, (G) LDL, (H) ApoB, (I) CA125, (J) CA199, and (K) HE-4. Each panel displays the discriminatory capability of the respective biomarker, with cutoff values selected based on the Youden index. Note: ROC, receiver operating characteristic; ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; HDL/LDL, high-/low-density lipoprotein; ApoB, apolipoprotein B; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4.

Analysis of risk factors for EC using logistic regression

Logistic regression was used to evaluate potential risk factors for EC. As shown in **Table 4**, fourteen variables were dichotomized, with continuous variables categorized based on optimal cutoff values from ROC curve analysis. Categorical variables were coded as 1 (“yes” or “irregular”) and 0 (“no” or “regular”), with the

Table 4. Variable assignment

Variable name	Variable type	Assignment content
ET	(X)	≥0.735=1, <0.735=0
FBG	(X)	≥5.575=1, <5.575=0
ALT	(X)	≥17.45=1, <17.45=0
AST	(X)	≥20.55=1, <20.55=0
TG	(X)	≥1.255=1, <1.255=0
HDL	(X)	≥1.405=1, <1.405=0
LDL	(X)	≥2.935=1, <2.935=0
ApoB	(X)	≥91.5=1, <91.5=0
CA125	(X)	≥21.09=1, <21.09=0
CA199	(X)	≥16.17=1, <16.17=0
HE-4	(X)	≥55.5=1, <55.5=0
AUB	(X)	Yes =1, no =0
Menstrual cycle regularity	(X)	Irregular =1, regular =0
Polycystic features	(X)	Yes =1, no =0
Sample type	(Y)	EC =1, control =0

Note: ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; HDL/LDL, high-/low-density lipoprotein; ApoB, apolipoprotein B; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4; AUB, abnormal uterine bleeding.

sample type coded as 1 for EC and 0 for controls. Univariate analysis identified statistically significant associations (P<0.05) between all examined variables and EC risk. Several biomarkers showed particularly strong associations, including HE-4 (OR=31.19, P<0.001), ET (OR=21.666, P<0.001), AST (OR=9.2, P<0.001), TG (OR=6.253, P<0.001), and CA199 (OR=6.246, P<0.001). Multivariate regression modeling revealed that most variables maintained independent predictive value after adjustment (P<0.05), except for LDL (OR=2.322, P=0.115) and menstrual cycle regularity (OR=1.771, P=0.408). The most robust independent predictors were HE-4 (OR=24.416, P<0.001), ET (OR=22.356, P<0.001), and TG (OR=13.794, P<0.001), as shown in **Table 5**.

Interaction analysis of EC with associated risk factors

The interaction analysis revealed significant associations between EC and several clinical and biochemical variables. The model intercept was -35.041 (P<0.001), indicating a strong baseline effect. Key variables significantly associated with EC risk included ET (Estimate =4.472, P<0.001), FBG (Estimate =0.691, P=0.040), AST (Estimate =0.327, P<0.001), TG (Estimate =3.609, P=0.001), CA125 (Estimate =0.195, P<0.001), CA199 (Estimate =0.252, P<0.001), HE-4 (Estimate =0.100, P<0.001), AUB (Estimate =2.149, P<0.001), and polycystic features (Estimate =3.193, P=0.006). No significant associations were observed for ALT (Estimate =0.079, P=0.199), HDL (Estimate =-0.658, P=0.349), or ApoB (Estimate =0.022, P=0.129). Details are in **Table 6** and **Figure 4**.

Nomogram model for predicting EC risk

The nomogram model, which incorporates 12 characteristic variables, provides a quantitative tool for individualized EC risk prediction. As illustrated in **Figure 5**, the nomogram displays variables in descending order of importance: HDL, AUB, CA199, CA125, FBG, ALT, ET, AST, TG, ApoB, and HE-4. Each variable is assigned a corresponding score on the scoring axis, and the cumulative score is used to predict EC risk probability. The model demonstrates differential weighting among predictors, with HE-4 and ET contributing most substantially to the total score, while HDL and TG exhibit more modest contributions. This suggests varying predictive weights for different variables in assessing EC risk (**Figure 5**).

Internal validation in the training set: model performance evaluation

Internal validation performed on the training set demonstrated exceptional predictive performance. ROC curve analysis yielded an AUC of 0.984 (**Figure 6A**), indicating excellent discriminative ability between the EC and control groups. Calibration, assessed via 1,000 bootstrap resamples, revealed excellent agreement between predicted probabilities and observed outcomes, with a mean absolute error (MAE) of 0.004, mean squared error (MSE) of 0.00003, and 0.9 quantile absolute error of 0.008. The GOF test showed no significant deviation (P=0.271), further supporting the model's excellent fit. Additional validation metrics confirm-

Table 5. Univariate and multivariate logistic regression analysis of risk factors for endometrial cancer

Variable name	Univariate			Multivariate		
	OR	P	95% CI	OR	P	95% CI
ET	21.666	<0.001	12.022-42.15	22.356	<0.001	7.234-84.243
FBG	3.091	<0.001	2.036-4.725	5.959	0.002	1.966-20.282
ALT	4.241	<0.001	2.786-6.544	7.55	<0.001	2.605-25.252
AST	9.2	<0.001	5.875-14.692	8.191	<0.001	3.033-24.497
TG	6.253	<0.001	3.984-9.973	13.794	<0.001	4.502-50.574
HDL	0.386	<0.001	0.251-0.591	0.312	0.042	0.097-0.935
LDL	2.673	<0.001	1.79-4.018	2.322	0.115	0.829-6.898
ApoB	2.176	<0.001	1.458-3.261	4.175	0.008	1.516-12.756
CA125	5.543	<0.001	3.524-8.923	5.284	0.003	1.845-16.516
CA199	6.246	<0.001	4.051-9.791	3.754	0.013	1.355-11.212
HE-4	31.19	<0.001	17.539-59.083	24.416	<0.001	8.341-85.142
AUB	4.984	<0.001	3.279-7.657	3.656	0.012	1.359-10.402
Menstrual cycle regularity	1.662	0.031	1.047-2.64	1.771	0.408	0.468-7.183
Polycystic features	5.079	<0.001	2.259-13.605	12.867	0.014	1.851-109.103

Note: OR, odds ratio; CI, confidence interval; ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; HDL/LDL, high-/low-density lipoprotein; ApoB, apolipoprotein B; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4; AUB, abnormal uterine bleeding.

Table 6. Interaction between endometrial cancer and risk factors

Variable name	Estimate	Std Error	Z value	Pr (>|z|)
(Intercept)	-35.041	5.379	-6.514	<0.001
ET	4.472	1.129	3.963	<0.001
FBG	0.691	0.336	2.054	0.040
ALT	0.079	0.061	1.284	0.199
AST	0.327	0.087	3.749	<0.001
TG	3.609	1.091	3.308	0.001
HDL	-0.658	0.703	-0.936	0.349
ApoB	0.022	0.014	1.519	0.129
CA125	0.195	0.053	3.656	<0.001
CA199	0.252	0.071	3.568	<0.001
HE4	0.100	0.018	5.621	<0.001
AUB	2.149	0.601	3.575	<0.001
Polycystic features	3.193	1.159	2.755	0.006

Note: ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; HDL, high-/low-density lipoprotein; ApoB, apolipoprotein B; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4; AUB, abnormal uterine bleeding.

ed strong calibration and discriminatory performance, with Dxy=0.9687, coefficient of determination (R^2) =0.8735, slope =0.9855, and maximum calibration error (Emax) =0.0039 (**Figure 6B**). Decision curve analysis (DCA) (**Figure 6C**) revealed substantial clinical net benefits across the entire risk threshold spectrum (0-99%), peaking at 39.90%, highlighting the model's clinical value.

Internal validation of the model performance on the validation set

Internal validation on the independent validation set confirmed the model's robust predictive performance. ROC curve analysis yielded an outstanding AUC of 0.987, demonstrating excellent discriminative ability in distinguishing EC from controls (**Figure 7A**). Calibration (based on 1,000 bootstrap samples) indicated excellent agreement between predicted probabilities and observed outcomes, with MAE of 0.014, MSE of 0.00038, and 0.9 quantile absolute error of 0.03. The GOF test further confirmed superior model fit (P=0.807), showing no significant deviation from ideal calibration. Other validation metrics included Somers' Dxy=0.9741, R^2=0.8810, calibration slope =0.9695, and Emax=0.0125, confirming exceptional calibration and discrimination (**Figure 7B**). DCA demonstrated strong clinical utility, with peak net benefit of 40.78% (**Figure 7C**), supporting its role in clinical decision-making.

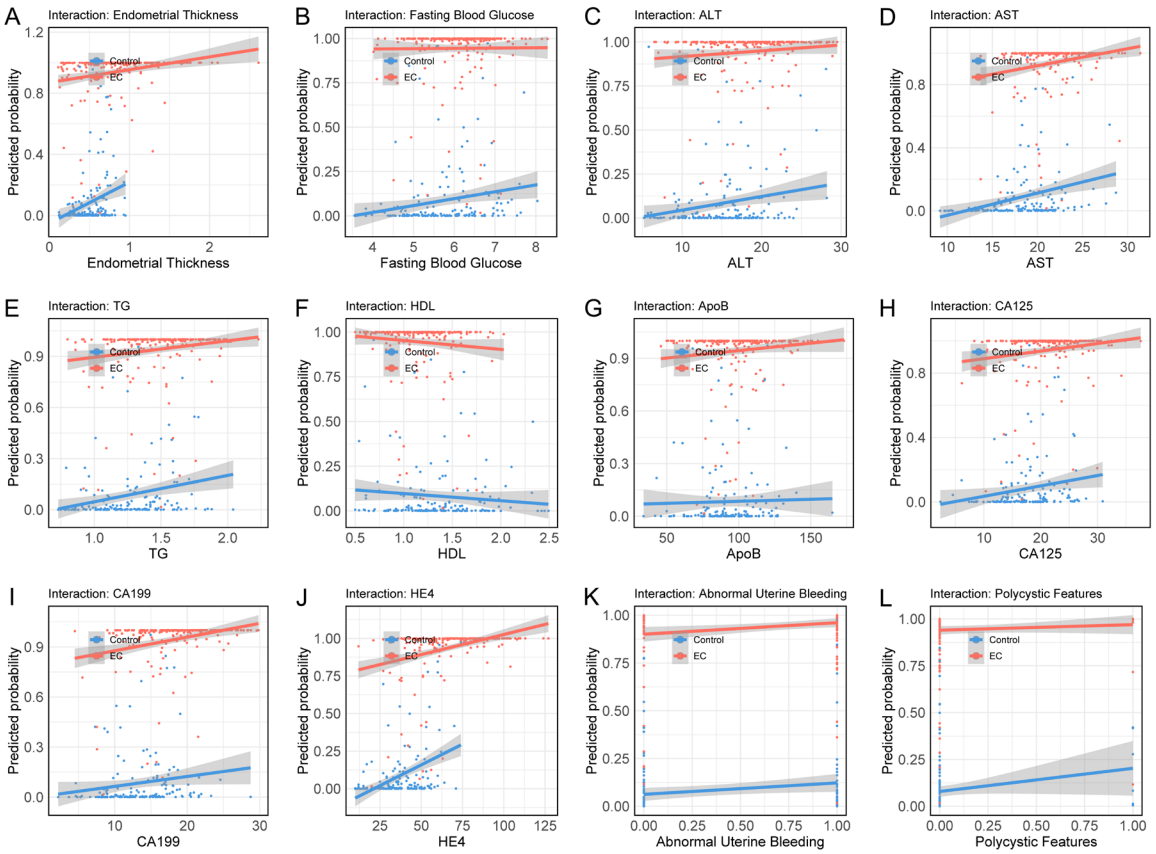

Figure 4. Interaction analysis of EC with associated risk factors. The predicted probability of EC in relation to: (A) ET, (B) FBG, (C) ALT, (D) AST, (E) TG, (F) HDL, (G) ApoB, (H) CA125, (I) CA19-9, (J) HE-4, (K) AUB, and (L) polycystic features. Note: EC, endometrial cancer; ET, endometrial thickness; FBG, fasting blood glucose; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; HDL, high-/low-density lipoprotein; ApoB, apolipoprotein B; CA125/199, cancer antigen 125/199; HE-4, human epididymis protein 4; AUB, abnormal uterine bleeding.

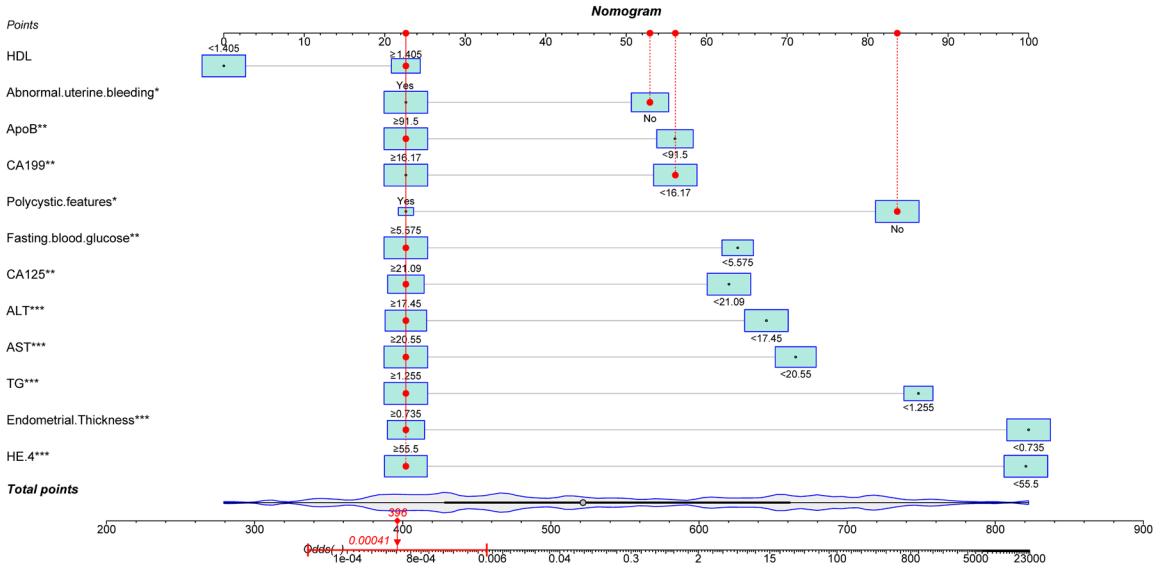

Figure 5. Nomogram for predicting endometrial cancer risk based on 12 characteristic variables. Note: HDL, high-density lipoprotein; AUB, abnormal uterine bleeding; ApoB, apolipoprotein B; CA199, cancer antigen 199; FBG, fasting blood glucose; CA125, cancer antigen 125; ALT, alanine aminotransferase; AST, aspartate aminotransferase; TG, triglycerides; ET, endometrial thickness; HE-4, human epididymis protein 4.

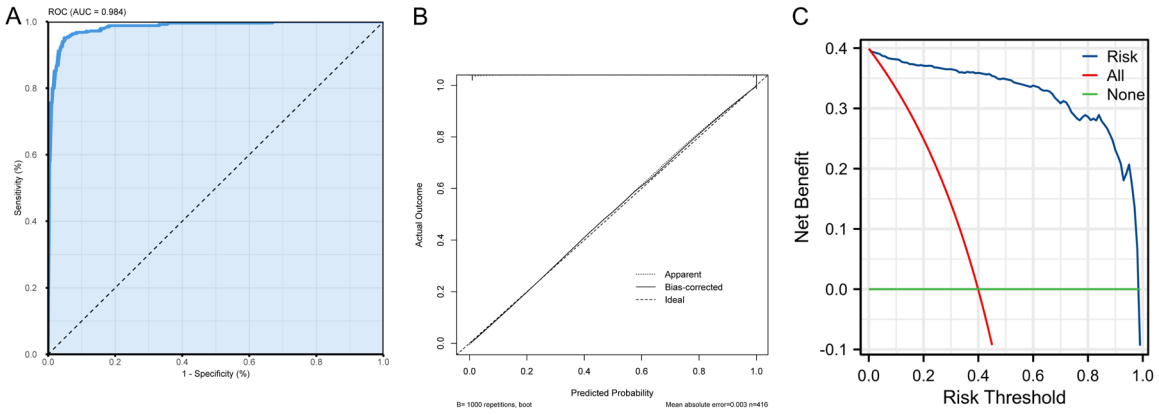

Figure 6. Internal validation in the training set: ROC curve, calibration curve, and DCA analysis. A. The ROC curve illustrates the model's discriminative ability, with an AUC of 0.984. The x-axis represents 1-specificity, while the y-axis corresponds to sensitivity. B. The calibration curve, generated using 1,000 bootstrap resamples, displays the relationship between predicted probabilities and actual outcomes. The blue line represents the fitted calibration curve, while the red dashed line indicates the ideal calibration line, demonstrating strong agreement between predictions and observations. C. The DCA compares the net benefit of the model (blue line) against the "treat-all" (red line) and "treat-none" (green line) strategies, highlighting the model's clinical value across different risk thresholds. Note: ROC, receiver operating characteristic; DCA, decision curve analysis; AUC, area under the curve.

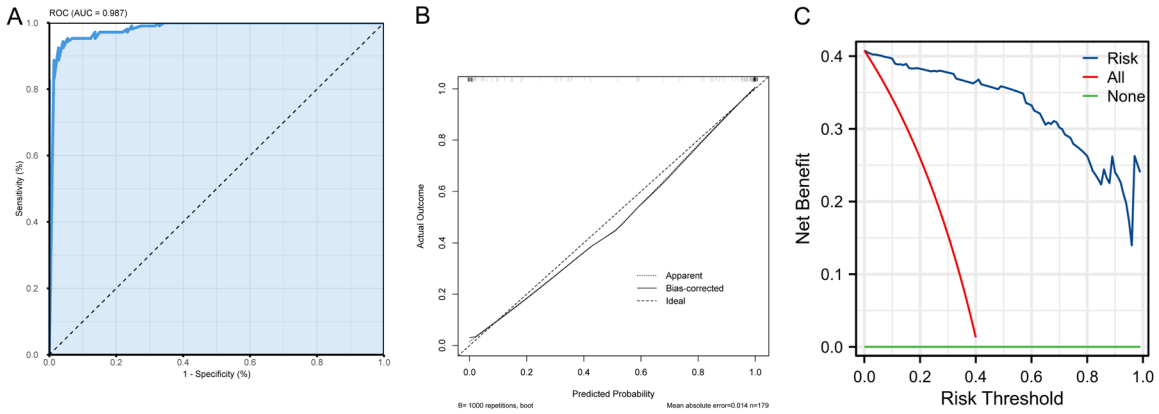

Figure 7. Internal validation: model performance assessment. A. ROC analysis showing exceptional diagnostic accuracy (AUC=0.987). Axes represent sensitivity versus 1-specificity. B. Bootstrap-corrected calibration plot (1000 iterations) comparing predicted probabilities (blue line) against ideal calibration (red dashed line). C. DCA quantifying clinical utility, with model performance (blue) superior to both "treat-all" (red) and "treat-none" (green) strategies. Note: ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis.

External validation of the predictive model: comprehensive performance assessment

The external validation on the independent external cohort confirmed the model's robust predictive capability. ROC analysis yielded an AUC of 0.964 (**Figure 8A**), demonstrating strong discriminative power between EC and control cases. Calibration using 1,000 bootstrap resamples revealed excellent agreement between predicted probabilities and observed outcomes, with MAE=0.011, MSE=0.00023, and 90th percentile absolute error of 0.025. The GOF test (P=0.241) indicated no significant miscalibration. Additional validation metrics reinforced the model's performance: Dxy=0.9287, R^2=0.7920, calibration slope =0.9739, and Emax=0.0078 (**Figure 8B**). DCA (**Figure 8C**) demonstrated superior clinical utility across the entire risk threshold spectrum

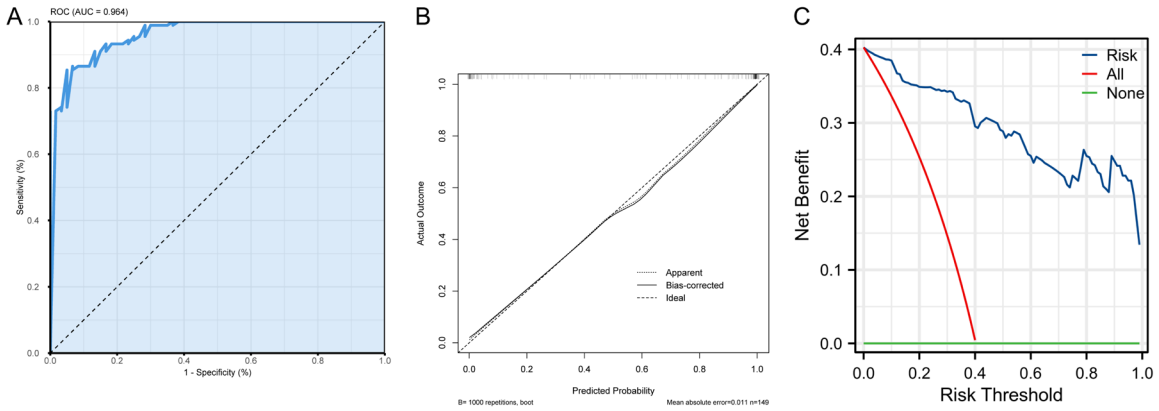

Figure 8. Performance evaluation of the externally validated prediction model. A. ROC Curve: The model demonstrates exceptional discriminative ability, with an AUC of 0.964. The x-axis represents 1-specificity, while the y-axis corresponds to sensitivity. B. Calibration Curve (Bootstrap Resampling: 1,000 iterations): The blue line depicts the agreement between predicted probabilities and observed outcomes, while the red dashed line indicates the ideal calibration reference. The close alignment confirms the model's high predictive accuracy. C. DCA: The net clinical benefit of the model is represented by the blue curve, compared against two extreme strategies: treat-all (red line) and treat-none (green line). Note: ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis.

(0-99%), with peak net benefit reaching 40.26%, significantly outperforming both "treat-all" and "treat-none" strategies.

Comparative analysis of the diagnostic performance between tumor markers and the comprehensive risk model

A comparative evaluation of the diagnostic utility of individual tumor markers (CA125, CA199, HE-4) and the composite risk model (Risk) across the training, validation, and external validation sets revealed that the comprehensive risk model consistently outperformed all individual biomarkers in all three datasets.

Training set analysis: The Risk model demonstrated significantly higher discriminatory accuracy than CA125 (AUC difference: -0.262; P<0.001), CA199 (AUC difference: -0.213; P<0.001), and HE-4 (AUC difference: -0.078; P<0.001). Among tumor markers, HE-4 significantly outperformed both CA125 (AUC difference: -0.184; P<0.001) and CA199 (AUC difference: -0.135; P<0.001).

Validation set analysis: In the validation set, the Risk model showed significantly greater diagnostic accuracy compared to CA125 (AUC difference: -0.231; P<0.001), CA199 (AUC difference: -0.189; P<0.001), and HE-4 (AUC difference: -0.068; P<0.001). HE-4 outperformed both CA125 (AUC difference: -0.162; P<0.001) and CA199 (AUC difference: -0.121; P<0.001).

External validation set analysis: In the external validation cohort, the Risk model maintained diagnostic superiority over CA125 (AUC difference: -0.233; P<0.001) and CA199 (AUC difference: -0.244; P<0.001). HE-4 also outperformed both CA125 (AUC difference: -0.218; P<0.001) and CA199 (AUC difference: -0.228; P<0.001). However, no statistically significant difference was observed between the Risk model and HE-4 (AUC difference: -0.015; P=0.435).

Collectively, these findings demonstrate that the comprehensive risk model achieves the highest diagnostic accuracy across all evaluated datasets (see **Table 7** and **Figure 9** for detailed comparisons).

Discussion

Among various hormone-related parameters, ET and PCOS-specific characteristics (such as hyperandrogenism and ovulatory dysfunction) are key determinants of EC risk [21, 22]. ET, a readily measurable imaging biomarker, reflects the proliferative status of the endometrial lining. Excessive and sustained proliferation in this tissue is a critical precursor to malignant transformation. A wealth of evidence links

Table 7. AUC comparisons between tumor markers and the risk model in the training, validation, and external validation sets

Marker1	Marker2	Z_value	*P*_value	AUC_difference	CI_lower_upper
CA125	CA199	-1.444	0.149	-0.049	-0.116 - 0.018
CA125	HE-4	-6.324	<0.001	-0.184	-0.241 - -0.127
CA125	Risk	-10.305	<0.001	-0.262	-0.312 - -0.212
CA199	HE-4	-4.99	<0.001	-0.135	-0.188 - -0.082
CA199	Risk	-9.202	<0.001	-0.213	-0.258 - -0.167
HE-4	Risk	-5.653	<0.001	-0.078	-0.105 - -0.051
CA125	CA199	-0.854	0.393	-0.041	-0.136 - 0.054
CA125	HE-4	-4.043	<0.001	-0.162	-0.241 - -0.084
CA125	Risk1	-6.497	<0.001	-0.231	-0.300 - -0.161
CA199	HE-4	-3.451	<0.001	-0.121	-0.190 - -0.052
CA199	Risk1	-5.901	<0.001	-0.189	-0.252 - -0.126
HE-4	Risk1	-3.721	<0.001	-0.068	-0.105 - -0.032
CA125	CA199	0.18	0.857	0.01	-0.103 - 0.124
CA125	HE-4	-4.689	<0.001	-0.218	-0.309 - -0.127
CA125	Risk2	-5.927	<0.001	-0.233	-0.310 - -0.156
CA199	HE-4	-4.903	<0.001	-0.228	-0.320 - -0.137
CA199	Risk2	-5.64	<0.001	-0.244	-0.328 - -0.159
HE-4	Risk2	-0.781	0.435	-0.015	-0.053 - 0.023

Note: AUC, area under the curve; CI, confidence interval; CA125, cancer antigen 125; CA199, cancer antigen 199; HE-4, human epididymis protein 4. Risk = training st, Risk1 = validation set, Risk2 = external validation set.

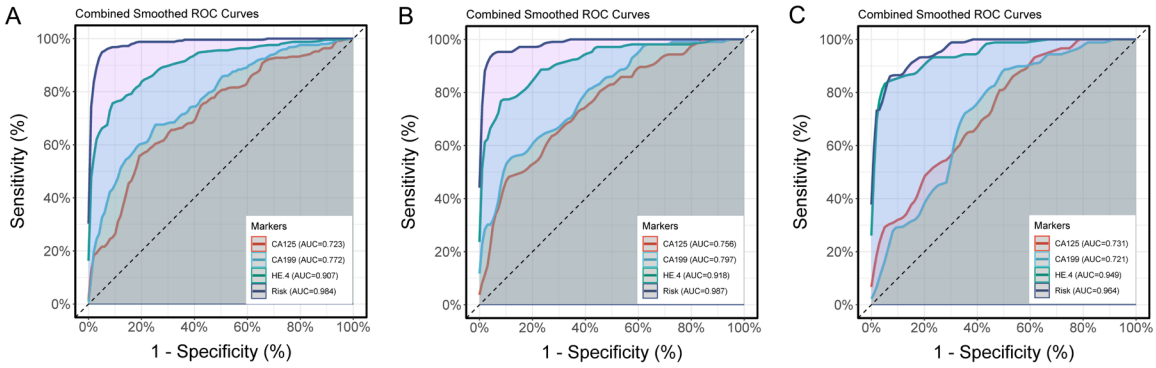

Figure 9. Smoothed ROC curves for the training set, validation set, and external validation set. A. Smoothed ROC curve for the training set. B. Smoothed ROC curve for the validation set. C. Smoothed ROC curve for the external validation set. Note: ROC, receiver operating characteristic; CA125, cancer antigen 125; AUC, area under the curve; CA199, cancer antigen 199; HE-4, human epididymis protein 4.

increased ET to heightened EC risk, particularly among PCOS patients who are not receiving hormonal interventions. In these individuals, unopposed estrogen exposure promotes endometrial hyperplasia, creating a favorable environment for carcinogenesis [23]. Therefore, ET measurement plays a pivotal role in early EC screening protocols, offering valuable diagnostic insights, especially in the PCOS population. Furthermore, PCOS is characterized by hyperandrogenism and chronic anovulation, which disrupt endocrine balance and lead to prolonged estrogen stimulation of the endometrium, thus increasing EC risk [24]. The common co-occurrence of PCOS with MetS exacerbates this risk by creating a harmful metabolic environment, further underscoring the need for close monitoring and proactive management of endocrine abnormalities during early screening initiatives.

Metabolic-related factors, including FBG, TG, HDL, LDL, and apolipoprotein B (ApoB), significantly influence EC risk [25]. MetS, strongly associated with insulin resistance, hyperglycemia, and dyslipidemia, collectively contributes to an elevated risk of EC. Shi et al. [13] demonstrated that metabolic disturbances such as obesity and diabetes markedly increase EC risk through insulin resistance, a key feature of MetS. Notably, elevated FBG levels are often coupled with hyperinsulinemia, which activates the IGF-1 pathway, stimulating abnormal endometrial cell proliferation and carcinogenesis [26, 27]. Thus, FBG not only serves as a biomarker of metabolic dysregulation in MetS but also as a crucial indicator for early EC screening. Additionally, other metabolic abnormalities - such as hypertriglyceridemia, reduced HDL, and elevated LDL - intensify chronic low-grade inflammation, further promoting endometrial cell proliferation and malignant transformation [28]. Studies have shown that PGD, a gene shared by both PCOS and EC, promotes endometrial cell proliferation through glucose metabolic pathways, mirroring the pathogenic mechanisms of MetS [29]. Elevated ApoB levels, associated with dyslipidemia and endocrine dysfunction, also exacerbate EC risk [30]. These findings highlight the importance of early identification and management of metabolic factors, particularly in individuals with MetS, as a promising strategy for reducing EC incidence.

Tumor markers, including HE-4, CA125, and CA199, play crucial roles in early EC detection. HE-4, in particular, has emerged as a highly promising biomarker, offering superior diagnostic sensitivity and specificity compared to traditional markers. Literature reports [31] indicate that HE-4 is significantly overexpressed in EC patients' serum, exhibiting high sensitivity and specificity. Combining HE-4 with other biomarkers significantly enhances diagnostic accuracy [32]. In our study, HE-4 demonstrated the highest AUC, underscoring its exceptional performance in distinguishing EC patients from controls. When integrated with clinical parameters such as ET and FBG, HE-4 significantly improved risk prediction accuracy. Cuesta-Guardiola et al. [33] found that serum HE-4 outperformed CA125 in EC diagnosis, as CA125 showed limited elevation in EC patients. Although CA125 and CA199 are elevated in certain EC subtypes, their standalone sensitivity remains limited. Supporting evidence [34] suggests that HE-4 and CA125 are the most promising serum biomarkers for EC diagnosis, with HE-4 being particularly suitable for inclusion in diagnostic algorithms. The comprehensive risk model, which integrates hormonal, metabolic, and tumor marker profiles, effectively addresses the limitations of single-marker approaches.

PCOS and MetS, as interrelated endocrine and metabolic disorders, often present clinically in overlapping populations. Their synergistic interaction may significantly contribute to endometrial carcinogenesis [35]. PCOS patients, exhibiting hyperandrogenic states, experience persistent endometrial proliferation, which can eventually progress to malignancy. Zhong et al. [17] identified obesity, prolonged menstrual cycles, and dyslipidemia as major risk factors for endometrial hyperplasia in PCOS populations, while demonstrating the protective effects of metformin and hormonal therapies. Giordano et al. [36] showed that BMI and insulin levels in PCOS patients disrupt the balance between endometrial proliferation and apoptosis, heightening carcinogenic risk. Insulin resistance, a hallmark of PCOS, exacerbates hyperglycemia and hyperinsulinemia, creating metabolic conditions that favor the development of MetS [37]. In MetS pathophysiology, insulin resistance and chronic inflammation constitute key pathways [38]. MetS can drive excessive endometrial hyperplasia through dysregulated IGF-1 pathways and prolonged pro-inflammatory cytokine release [39]. This bidirectional PCOS-MetS interaction establishes a vicious cycle that substantially amplifies EC risk. Our interaction analysis confirms that the synergy between PCOS and MetS significantly influences EC development, highlighting the prominence of PCOS characteristics, AUB, FBG levels, and HE-4 in high-risk populations.

The nomogram model provides clinicians with a powerful risk assessment tool by quantitatively integrating predictive variables for intuitive individual risk evaluation in EC. Key variables such as HE-4 and ET exhibit substantial predictive value, highlighting their critical role in early screening. Kuai et al. [18] demonstrated that nomograms incorporating PCOS and ET achieved AUC of 0.889-0.956 for predicting endometrial hyperplasia/EC risk in young women. Li et al. [40] reported that nomograms combining

PCOS and metabolic factors yielded an AUC of 0.905 in premenopausal women. HE-4 emerged as a superior tumor marker with an AUC of 0.907, significantly outperforming conventional biomarkers such as CA125 and CA199. Our model showed high performance across training, validation, and external validation sets. Internal validation achieved an impressive AUC of 0.987, while external validation yielded an AUC of 0.964, demonstrating robust discriminative ability. Calibration curves revealed excellent agreement between predicted and observed outcomes, and DCA confirmed substantial clinical net benefits across a wide range of risk thresholds. Existing literature [41] supports the notion that nomogram models incorporating metabolic factors significantly improve EC diagnostic efficiency by optimizing patient selection for hysteroscopic examination. Compared to traditional single-marker approaches, the nomogram demonstrated superior performance in all validation sets, particularly in sensitivity and specificity for early-stage detection. This comprehensive risk model enhances diagnostic precision for identifying high-risk populations and facilitates personalized treatment strategies, improving clinical outcomes and survival rates.

While this study demonstrated promising results, several limitations must be acknowledged. First, the retrospective design and reliance on data from a single institution may limit the generalizability of our findings. Future multicenter prospective studies with larger cohorts are needed to validate the model's performance across diverse populations. Second, the analysis did not comprehensively account for other potential risk factors, such as genetic predisposition and environmental exposures, which may significantly contribute to endometrial carcinogenesis. Future research should broaden the scope to include these and other relevant variables to improve the model's predictive accuracy and clinical utility. Lastly, although the nomogram performed well in our cohort, its real-world clinical application requires further evaluation. In resource-constrained settings, future studies should focus on optimizing the integration of this model with existing screening protocols to develop cost-effective clinical decision-making strategies.

This study established a high-risk prediction model for EC by integrating the synergistic effects of PCOS and metabolic syndrome (MetS). The model demonstrated robust predictive accuracy and consistent performance across the training, validation, and external validation cohorts. Its ability to reliably identify high-risk individuals provides a promising tool for enhancing early detection and screening strategies for EC.

Disclosure of conflict of interest

None.

Address correspondence to: Yabin Guo, Center of Reproductive Medicine, Changde Hospital, Xiangya School of Medicine, Central South University (The First People's Hospital of Changde City), No. 818 Renmin Road, Changde 415000, Hunan, China. Tel: +86-0736-7788328; E-mail: gybhzh@163.com

References

[1] Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME and Singh N. Endometrial cancer. Lancet 2022; 399: 1412-1428.

[2] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-263.

[3] Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W and He J. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent 2024; 4: 47-53.

[4] Stener-Victorin E, Teede H, Norman RJ, Legro R, Goodarzi MO, Dokras A, Laven J, Hoeger K and Piltonen TT. Polycystic ovary syndrome. Nat Rev Dis Primers 2024; 10: 27.

[5] Ducote M, Schauer T, Ross R, Boyer LM, Stagg MP, Domangue E, Graham B, Garcia J, Stillwell C, Drews KL, Schauer PR, Cook MW, Jernigan A and Albaugh VL. High prevalence of dysfunctional uterine bleeding in candidates for metabolic/bariatric surgery: increased endometrial cancer risk? Surg Obes Relat Dis 2024; 20: 1172-1178.

[6] Santander N, Figueroa EG, González-Candia A, Maliqueo M, Echiburú B, Crisosto N and Salas-Pérez F. Oxidative stress in polycystic ovary syndrome: impact of combined oral contraceptives. Antioxidants (Basel) 2024; 13: 1168.

[7] Jungari M, Choudhary A and Gill NK. Comprehensive management of polycystic ovary syndrome: effect of pharmacotherapy, lifestyle modification, and enhanced adherence counseling. Cureus 2023; 15: e35415.

[8] Ignatov A and Ortmann O. Endocrine risk factors of endometrial cancer: polycystic ovary

syndrome, oral contraceptives, infertility, tamoxifen. Cancers (Basel) 2020; 12: 1766.
[9] Frandsen CLB, Gottschau M, Nøhr B, Viuff JH, Maltesen T, Kjær SK, Jensen A and Svendsen PF. Polycystic ovary syndrome and endometrial cancer risk: results from a nationwide cohort study. Am J Epidemiol 2024; 193: 1399-1406.
[10] Palomba S, Piltonen TT and Giudice LC. Endometrial function in women with polycystic ovary syndrome: a comprehensive review. Hum Reprod Update 2021; 27: 584-618.
[11] Richman I and Rittenberg E. Lifestyle modification for obesity management-a cornerstone and not a roadblock. JAMA Intern Med 2025; 185: 257.
[12] Butragueño J and Ruiz JR. Metabolic alliance: pharmacotherapy and exercise management of obesity. Nat Rev Endocrinol 2024; 20: 505-506.
[13] Shi J, Kraft P, Rosner BA, Benavente Y, Black A, Brinton LA, Chen C, Clarke MA, Cook LS, Costas L, Dal Maso L, Freudenheim JL, Frias-Gomez J, Friedenreich CM, Garcia-Closas M, Goodman MT, Johnson L, La Vecchia C, Levi F, Lissowska J, Lu L, McCann SE, Moysich KB, Negri E, O'Connell K, Parazzini F, Petruzella S, Polesel J, Ponte J, Rebbeck TR, Reynolds P, Ricceri F, Risch HA, Sacerdote C, Setiawan VW, Shu XO, Spurdle AB, Trabert B, Webb PM, Wentzensen N, Wilkens LR, Xu WH, Yang HP, Yu H, Du M and De Vivo I. Risk prediction models for endometrial cancer: development and validation in an international consortium. J Natl Cancer Inst 2023; 115: 552-559.
[14] Abdul Hafizz AMH, Mohd Mokthar N, Md Zin RR, P Mongan N, Mamat Yusof MN, Kampan NC, Chew KT and Shafiee MN. Insulin-like Growth Factor 1 (IGF1) and its isoforms: insights into the mechanisms of endometrial cancer. Cancers (Basel) 2025; 17: 129.
[15] Di Lorenzo M, Cacciapuoti N, Lonardo MS, Nasti G, Gautiero C, Belfiore A, Guida B and Chiurazzi M. Pathophysiology and nutritional approaches in polycystic ovary syndrome (PCOS): a comprehensive review. Curr Nutr Rep 2023; 12: 527-544.
[16] DiNicolantonio JJ and H O'Keefe J. Myo-inositol for insulin resistance, metabolic syndrome, polycystic ovary syndrome and gestational diabetes. Open Heart 2022; 9: e001989.
[17] Zhong X, Li Y, Liang W, Hu Q, Zeng A, Ding M, Chen D and Xie M. Clinical and metabolic characteristics of endometrial lesions in polycystic ovary syndrome at reproductive age. BMC Womens Health 2023; 23: 236.
[18] Kuai D, Tang Q, Tian W and Zhang H. Rapid identification of endometrial hyperplasia and endometrial endometrioid cancer in young women. Discov Oncol 2023; 14: 121.
[19] Berek JS, Matias-Guiu X, Creutzberg C, Fotopoulou C, Gaffney D, Kehoe S, Lindemann K, Mutch D and Concin N; Endometrial Cancer Staging Subcommittee, FIGO Women's Cancer Committee. FIGO staging of endometrial cancer: 2023. Int J Gynaecol Obstet 2023; 162: 383-394.
[20] Siddiqui S, Mateen S, Ahmad R and Moin S. A brief insight into the etiology, genetics, and immunology of polycystic ovarian syndrome (PCOS). J Assist Reprod Genet 2022; 39: 2439-2473.
[21] Swaroop M, Nevels A, Hubregsen M and Darby JP. Endometrial cancer of both uterine horns in a premenopausal patient with uterine didelphys: surgical approach and considerations for adjuvant treatment. Gynecol Oncol Rep 2024; 55: 101492.
[22] Wang SC, Wu CH, Fu HC, Ou YC, Tsai CC, Chen YY, Wang YW, Hunag SW, Huang SY, Lan J and Lin H. Estrogen/progesterone receptor expression and cancer antigen 125 level as preoperative predictors to estimate lymph node metastasis in endometrioid endometrial cancer. Int J Gynecol Pathol 2024; 43: 316-325.
[23] Thoprasert P, Phaliwong P, Smanchat B, Prommas S, Bhamarapravatana K and Suwannarurk K. Endometrial thickness measurement as predictor of endometrial hyperplasia and cancer in perimenopausal uterine bleeding: cross-sectional study. Asian Pac J Cancer Prev 2023; 24: 693-699.
[24] Doll KM, Pike M, Alson J, Williams P, Carey E, Stürmer T, Wood M, Marsh EE, Katz R and Robinson WR. Endometrial thickness as diagnostic triage for endometrial cancer among black individuals. JAMA Oncol 2024; 10: 1068-1076.
[25] Lin Q, Lu Y, Lu R, Chen Y, Wang L, Lu J and Ye X. Assessing metabolic risk factors for LVSI in endometrial cancer: a cross-sectional study. Ther Clin Risk Manag 2022; 18: 789-798.
[26] Kokts-Porietis RL, McNeil J, Nelson G, Courneya KS, Cook LS and Friedenreich CM. Prospective cohort study of metabolic syndrome and endometrial cancer survival. Gynecol Oncol 2020; 158: 727-733.
[27] Knapp P, Chabowski A, Harasiuk D and Górski J. Reversed glucose and fatty acids transporter expression in human endometrial cancer. Horm Metab Res 2012; 44: 436-441.
[28] Yeu TH, Omar IS, Sani SFA, Pathmanathan D, Goh BT, Ravindran N, Teo IH, Gunasagran Y, Mat Adenan NA, Chung I and Abd Jamil AH. Distinct lipid phenotype of cancer-associated fibroblasts (CAFs) isolated from overweight/obese endometrial cancer patients as assessed using raman spectroscopy. Appl Spectrosc 2023; 77: 723-733.

[29] Chen JM, Chen WH, Wang ZY, Zhou LY, Lin QY, Huang QY, Zheng LT, You HJ, Lin S and Shi QY. PGD: shared gene linking polycystic ovary syndrome and endometrial cancer, influencing proliferation and migration through glycometabolism. Cancer Sci 2024; 115: 2908-2922.
[30] Yang Z, Chen J, Wen M, Lei J, Zeng M, Li S, Long Y, Zhou Z and Wang C. Genetic association of lipids and lipid-lowering drug target genes with Endometrial carcinoma: a drug target Mendelian randomization study. Front Endocrinol (Lausanne) 2024; 15: 1446457.
[31] Behrouzi R, Barr CE and Crosbie EJ. HE4 as a biomarker for endometrial cancer. Cancers (Basel) 2021; 13: 4764.
[32] Liu J, Han L and Jiao Z. The diagnostic value of human epididymis protein 4 for endometrial cancer is moderate. Sci Rep 2021; 11: 575.
[33] Cuesta-Guardiola T, Carretero AQ, Martinez-Martinez J, Cuñarro-López Y, Pereira-Sánchez A, Fernández-Corona A and de Leon-Luis JA. Identification and characterization of endometrial carcinoma with tumor markers HE4 and CA125 in serum and endometrial tissue samples. J Turk Ger Gynecol Assoc 2021; 22: 161-167.
[34] Koleva V, Shefket S and Stoencheva S. Clinical significance of human epididymis protein 4 (HE4), cancer antigen 125 (CA125), the risk of ovarian malignancy algorithm (ROMA), and Copenhagen index (CPH-I) for the diagnosis of endometrial carcinoma. Folia Med (Plovdiv) 2025; 67.
[35] Helvaci N and Yildiz BO. Polycystic ovary syndrome as a metabolic disease. Nat Rev Endocrinol 2025; 21: 230-244.
[36] Giordano LA, Giordano MV, Célia Teixeira Gomes R, Dos Santos Simões R, Baracat MCP, Giordano MG, Ferreira-Filho ES, de Medeiros SF, Baracat EC and Soares-Júnior JM. Effects of clinical and metabolic variables and hormones on the expression of immune protein biomarkers in the endometrium of women with polycystic ovary syndrome and normal-cycling controls. Gynecol Endocrinol 2022; 38: 508-515.
[37] Saadati S, Mason T, Godini R, Vanky E, Teede H and Mousa A. Metformin use in women with polycystic ovary syndrome (PCOS): opportunities, benefits, and clinical challenges. Diabetes Obes Metab 2025; 27 Suppl 3: 31-47.
[38] Fu J, Zhang Y, Cai X and Huang Y. Predicting metformin efficacy in improving insulin sensitivity among women with polycystic ovary syndrome and insulin resistance: a machine learning study. Endocr Pract 2024; 30: 1023-1030.
[39] Nwabo Kamdje AH, Seke Etet PF, Kipanyula MJ, Vecchio L, Tagne Simo R, Njamnshi AK, Lukong KE and Mimche PN. Insulin-like growth factor-1 signaling in the tumor microenvironment: carcinogenesis, cancer drug resistance, and therapeutic potential. Front Endocrinol (Lausanne) 2022; 13: 927390.
[40] Li Z, Yin J, Liu Y and Zeng F. A risk prediction model for endometrial hyperplasia/endometrial carcinoma in premenopausal women. Sci Rep 2025; 15: 1019.
[41] Bottura B, Haddad RF, Alvarenga-Bezerra V, Campos V, Perez L, Resende C, Asencio FA, Liao AW, Gomes MTV, Zlotnik E and Moretti-Marques R. Developing a nomogram for prioritizing hysteroscopy in endometrial cancer diagnosis: a case-control study. J Clin Med 2024; 13: 1145.