

## Original Article

# Clinical study of classification and regression tree assisted endometrial biopsy in the diagnosis of endometrial carcinoma

Wei Fu, Min Zhang, Yuxin Zhou, Shujun Zhang, Lei Dong, Liying Wang

*Department of Gynecology, Changde Hospital, Xiangya School of Medicine, Central South University, Changde 415000, Hunan, China*

Received September 11, 2023; Accepted October 25, 2023; Epub November 15, 2023; Published November 30, 2023

**Abstract:** The early diagnosis of endometrial carcinoma is critical for improving patient survival and prognosis. However, the diagnostic efficiency of a single examination is often insufficient, because it is easy to cause misdiagnosis and missed diagnoses. Therefore, this study used the classification and regression tree (CART) algorithm to establish and validate a CART model to distinguish endometrial carcinoma from other endometrial lesions. The clinical data of 297 patients treated at Changde Hospital, Xiangya School of Medicine, Central South University between April 2021 and April 2023 for postmenopausal uterine effusion, postmenopausal vaginal bleeding, abnormal uterine bleeding and endometrial thickening were retrospectively analyzed. Among them, there were 203 cases of endometrial carcinoma and 94 cases of endometrial lesions. The pathological results from endometrial biopsy and hysteroscopic curettage were compared. The coincidence rate of endometrial biopsy was 90.34% (187/207) and the AUC, sensitivity, and specificity of the diagnosis of endometrial carcinoma were 0.920, 0.914, and 0.925, respectively. Six serological indicators with diagnostic significance were screened out: carbohydrate antigen 125 (CA125), carbohydrate antigen 19-9 (CA19-9), human epididymis secretory protein 4 (HE4), vascular endothelial growth factor (VEGF), D-dimer, and absolute neutrophil count (N). The AUC, sensitivity and specificity of the CART model based on the above indicators were 0.949, 0.979, and 0.896, respectively. The CART model is an intuitive and simple tool for the clinical diagnosis of endometrial carcinoma and endometrial lesions.

**Keywords:** Endometrial biopsy, endometrial carcinoma, decision tree, diagnosis

## Introduction

Endometrial carcinoma (EC), a prevalent malignant tumor in gynecology, originates from endometrial epithelial cells, particularly in postmenopausal women over 50 years old. The prognosis of EC is not ideal, and the age of onset has a significantly younger trend in recent years [1]. Early EC lacks typical symptoms, often manifested as irregular vaginal bleeding, so it is easy to be overlooked in clinic, resulting in delayed diagnosis and treatment [2]. Therefore, the exploration of early screening and late monitoring indicators of EC has gradually received attention. At present, laparoscopy or laparotomy combined with postoperative pathological testing is considered the gold standard for diagnosing EC. However, these methods are cumbersome and costly, thus not

suitable for large-scale population screening [3]. Previous research has revealed that a one-time uterine tissue aspiration tube for endometrial biopsy has the advantages of shorter operation time, less invasive, no need to dilate the cervix, less pain, and low cost. This method can be widely used in screening endometrial diseases [4]. Serological markers are a group of small molecules that are stable in the blood, with a characteristic of easy detection, making them ideal markers for cancer screening [5]. For instance, CA125, CA19-9, and HE4 are used as auxiliary diagnostic indicators for EC in clinical practice [6].

Classification and regression tree (CART) algorithm, as a machine learning algorithm, is a product of the current era of network intelligent diagnosis and treatment technology. The CART

# Diagnosis of endometrial carcinoma

algorithm is not limited by fixed modeling rules, and can better deal with massive disordered data. Compared with the traditional Logistic regression model, CART has advantages, such as easy interpretation of the output results, which can help doctors make more accurate disease judgments [7]. At present, there is no clear distinction in the clinical application value of relevant serological indicators in the diagnosis of EC. In addition, the use of a single diagnostic method may lead to misdiagnosis or missed diagnoses [8, 9]. Combining the results of endometrial biopsy with related indicators to develop a model for joint diagnosis may provide a comprehensive evaluation of patient conditions and improve diagnostic accuracy. But no relevant research on this aspect has been found. Therefore, this work constructed a CART model using results of endometrial biopsy and serological indicators to distinguish EC from other endometrial lesions, with a goal of providing a basis for early diagnosis of EC.

## Materials and methods

### *Case selection and ethics approval*

Patients treated at Changde Hospital, Xiangya School of Medicine, Central South University between April 2021 and April 2023 for postmenopausal uterine effusion, postmenopausal vaginal bleeding, abnormal uterine bleeding, and endometrial thickening were selected as subjects. Inclusion criteria: (1) Patients who received both endometrial biopsy via one-time uterine tissue aspiration and hysteroscopic curettage in our hospital and obtained pathological results; (2) Patients with comprehensive clinical information. Exclusion criteria: (1) Patients with vaginitis, cervical polyps, and other malignant tumors; (2) Patients who received hormone therapy within 1 year; (3) Patients with abnormal coagulation function or heart, lung, liver, or renal insufficiency. There were 297 patients meeting the mentioned criteria, and they were categorized into two groups based on their pathological diagnosis results from hysteroscopic curettage: the EC group (n=203) and the endometrial lesion group (n=94). All patients provided an informed consent for participating in the treatment. This study was approved by the Ethics Committee of Changde Hospital, Xiangya School of Medicine, Central South University. The flow chart of this study is shown in **Figure 1**.

### *Methods of sampling and diagnosis*

To collect the endometrial tissue, the patients first underwent endometrial biopsy using a disposable uterine tissue suction tube. The uterine tissue suction tube was slowly inserted into the uterine fundus, and the endometrial tissue was sucked out and placed in a container containing formaldehyde aqueous solution (specimen 1). Hysteroscopy was performed to observe the endometrial sampling under hysteroscope. Curettage was also performed to obtain endometrial specimens (specimen 2). The results of hysteroscopic curettage were used as the golden criteria [1]. Specimens 1 and 2 were pathologically diagnosed by two doctors respectively. The pathological results were divided into (1) endocrine changes in the endometrium (including secretory endometrium, proliferative endometrium, and menstrual endometrium); (2) endometrial hyperplasia (including simple, complicated, and atypical hyperplasia); (3) endometrial polyps; (4) endometrial carcinoma.

### *Data collection*

The clinical data of patients were collected from their electronic medical records, encompassing details such as age, body mass index (BMI), carbohydrate antigen 125 (CA125), carbohydrate antigen 19-9 (CA19-9), vascular endothelial growth factor (VEGF), human epididymis secretory protein 4 (HE4), D-dimer, red blood cell count (RBC), white blood cell count (WBC), absolute neutrophil count (N), absolute lymphocyte count (L), monocyte count (MON), mean corpuscular hemoglobin content (MCH), mean corpuscular volume (MCV), red blood cell distribution width (RDW), platelet count (PLT), mean platelet volume (MPV), platelet distribution width (PDW), neutrophil to lymphocyte ratio (NLR), monocyte to lymphocyte ratio (MLR), and platelet to lymphocyte ratio (PLR). All serological indicators were measured in fasting blood on the next day after admission.

### *Statistical analysis*

Statistical analyses were performed using SPSS 26.0 for Windows. Data with a normal distribution were reported as mean  $\pm$  standard deviation, and t-test was performed to compare the group data. The non-normally distrib-

## Diagnosis of endometrial carcinoma

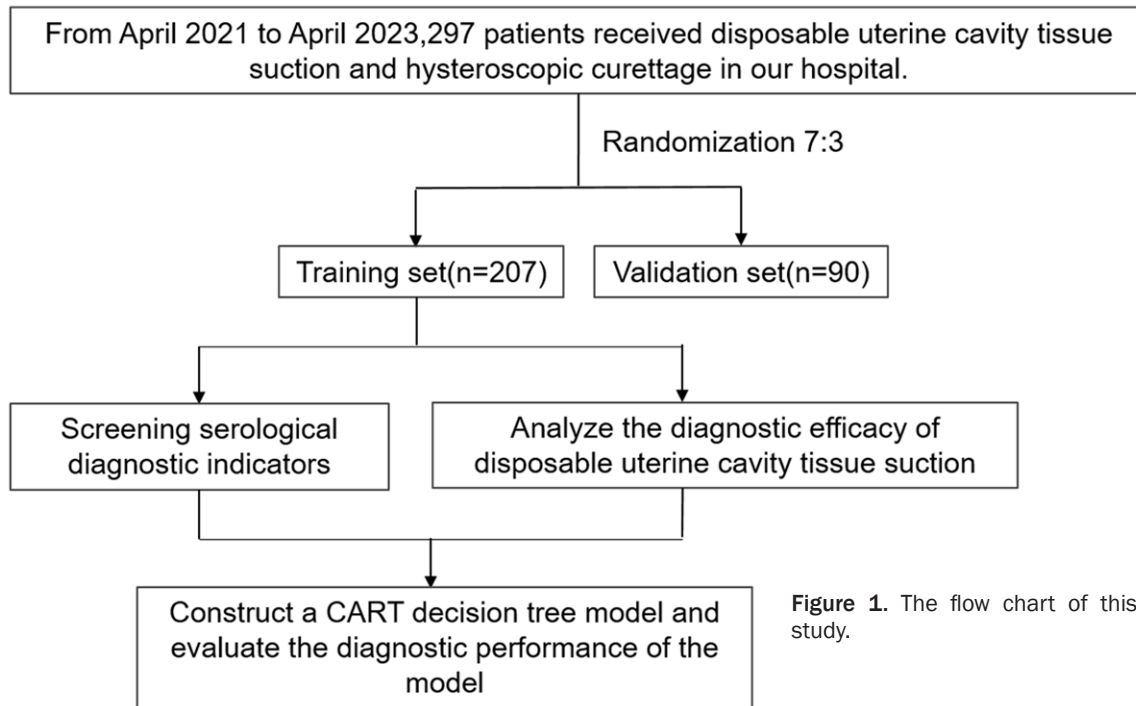


Figure 1. The flow chart of this study.

uted measurement data were reported as median (quartile), and Mann-Whitney U test was employed to compare these data. Count data were expressed as a percentage of the total number of instances. R 4.2.3 software was utilized for LASSO regression analysis and multivariate logistic regression analysis. The Wayne diagram was made to screen out the indicators with statistical significance in LASSO regression and logistic regression as effective diagnostic indicators. The data were randomly divided into a training set and a verification set in a 7:3 ratio. The CART model was built with “rpart” package in R, with the main parameters set to `parms=list(split="gini")` and `method="class"`. The diagnostic efficiency was verified by the ROC curve, calibration curve, area under the ROC curve (AUC), sensitivity and specificity. The delong test was used to compare the AUC value.  $P < 0.05$  was set as the statistical significance level.

### Results

#### *Comparison of data between the training set and the validation set*

The clinical data of 297 patients were collected and randomly divided into a training set (n=207) and a validation set (n=90) according to the ratio of 7:3. There was no significant dif-

ference in age, BMI and serological indexes between the two cohorts (Table 1).

#### *The diagnostic efficacy of endometrial biopsy for endometrial lesions*

The sensitivity, specificity, and AUC of the pathological results of endometrial biopsy were analyzed. It was found that the coincidence rate of pathological diagnosis of endometrial biopsy was 90.34% (187/207). The sensitivity of the diagnosis of EC was 0.914, the sensitivity of the diagnosis of endometrial endocrine changes was 0.941, the sensitivity of the diagnosis of endometrial hyperplasia was 0.927, and the sensitivity of the diagnosis of endometrial polyps was 0.556 (Table 2).

#### *Clinical data comparison between the two patient groups*

A comparison was conducted on the clinical data of the two groups. The findings revealed notable variations in CA125, CA19-9, HE4, VEGF, D-dimer, MPV, N, and NLR between the two groups (Table 3).

#### *LASSO regression analysis*

After univariate analysis, the regularization method of LASSO regression was used to screen the diagnostic indicators (Figure 2A).

## Diagnosis of endometrial carcinoma

**Table 1.** Comparison of data between the training set and the validation set [mean  $\pm$  SD/M ( $P_{25}$ ,  $P_{75}$ )]

Patient characteristics	Training set (n=207)	Validation set (n=90)	t/Z	P
Age, year	51.34 $\pm$ 8.09	52.39 $\pm$ 8.09	1.025	0.306
BMI, kg/m <sup>2</sup>	23.74 $\pm$ 2.57	23.67 $\pm$ 2.97	0.222	0.825
CA125, U/mL	27.02 (15.71, 42.82)	28.18 (19.77, 44.18)	0.835	0.404
CA19-9, U/mL	20.84 (11.84, 32.32)	21.50 (13.01, 35.01)	0.948	0.343
HE4, pmol/L	70.20 (48.20, 96.10)	70.30 (52.50, 88.00)	0.209	0.835
VEGF, ng/L	291.68 $\pm$ 95.87	291.56 $\pm$ 102.00	0.009	0.993
D-dimer, ng/mL	351.40 $\pm$ 143.79	336.00 $\pm$ 133.19	0.864	0.388
RBC, 10 <sup>9</sup> /L	3.96 $\pm$ 1.28	3.82 $\pm$ 1.14	0.860	0.391
MCH, g/L	128.44 $\pm$ 1.74	128.26 $\pm$ 1.88	0.814	0.416
MCV, fL	87.40 $\pm$ 1.32	87.22 $\pm$ 1.27	1.115	0.266
RDW, fL	44.43 $\pm$ 2.99	44.23 $\pm$ 2.50	0.540	0.590
PLT, 10 <sup>9</sup> /L	223.97 $\pm$ 63.02	230.28 $\pm$ 72.78	0.753	0.452
MPV, fL	12.10 (10.80, 12.90)	11.75 (10.20, 12.90)	1.429	0.153
PDW, fL	13.96 $\pm$ 2.09	13.55 $\pm$ 2.19	1.525	0.128
WBC, 10 <sup>9</sup> /L	6.52 $\pm$ 1.42	6.26 $\pm$ 1.31	0.860	0.391
N, 10 <sup>9</sup> /L	3.76 $\pm$ 1.27	3.77 $\pm$ 1.27	1.462	0.145
L, 10 <sup>9</sup> /L	1.90 (1.52, 2.36)	1.96 (1.53, 2.27)	0.220	0.826
MON, 10 <sup>9</sup> /L	0.33 $\pm$ 0.11	0.34 $\pm$ 0.11	0.662	0.508
NLR	1.95 (1.47, 2.53)	1.94 (1.36, 2.52)	0.078	0.938
PLR	111.71 (88.02, 150.73)	114.43 (91.56, 159.46)	0.561	0.575
MLR	0.17 (0.12, 0.23)	0.17 (0.14, 0.22)	0.357	0.721

Note: BMI: body mass index; CA125: carbohydrate antigen 125; CA19-9: carbohydrate antigen 19-9; VEGF: vascular endothelial growth factor; HE4: human epididymis secretory protein 4; RBC: red blood cell count; MCH: mean corpuscular hemoglobin content; MCV: mean corpuscular volume; RDW: red blood cell distribution width; PLT: platelet count; MPV: mean platelet volume; PDW: platelet distribution width; WBC: white blood cell count; N: absolute neutrophil count; L: absolute lymphocyte count; MON: monocyte count; NLR: neutrophil to lymphocyte ratio; PLR: platelet to lymphocyte ratio; MLR: monocyte to lymphocyte ratio.

The ideal input value was determined using the 10-fold cross-validation approach (**Figure 2B**). In the figure, the two dashed lines correspond to lambda.min and lambda.lse. The former signifies the  $\lambda$  value at which the mean square error is minimized, while the latter corresponds to the  $\lambda$  value associated with the smallest mean square error considering standard error. The model opts for the  $\lambda$  value of 0.006, which aligns with lambda.min. Ultimately, CA125, CA19-9, HE4, VEGF, D-dimer, N, and MPV were the selected variables.

### Multivariate logistic regression analysis

CA125 (the original value), CA19-9 (the original value), HE4 (the original value), VEGF (the original value), D-dimer (the original value), N (the original value), NLR (the original value) and MPV (the original value) were used as independent variables. The dependent variable utilized was the diagnostic outcomes of hysterosco-

pic curettage, with a value of 1 indicating EC and 0 denoting endometrial lesions. The results of the multivariate logistic regression analysis indicated that CA125, CA19-9, HE4, VEGF, D-dimer, and N were significant factors influencing the differentiation between patients with EC and those with endometrial lesions (**Table 4**).

### Serological indicators for the diagnosis of EC

The diagnostic indicators screened by LASSO regression and Logistic regression were intersected, and six common indicators, namely CA125, CA19-9, HE4, VEGF, D-dimer, and N, were obtained to make a Wayne diagram, as shown in **Figure 3**.

### Comparison of the efficacy of serological indicators in the diagnosis of EC

By comparing the AUC values of each diagnostic index, it was found that the diagnostic effi-

## Diagnosis of endometrial carcinoma

**Table 2.** Diagnostic efficacy of endometrial biopsy for endometrial lesions (n=207)

Endometrial biopsy	Hysteroscopic curettage							
	Endocrine changes in endometrium		Endometrial hyperplasia		Endometrial polyps		Endometrial carcinoma	
	Yes	No	Yes	No	Yes	No	Yes	No
Yes	16	5	38	8	5	2	128	5
No	1	185	3	158	4	196	12	62
Sensitivity	0.941		0.927		0.556		0.914	
Specificity	0.974		0.952		0.990		0.925	
AUC	0.957		0.939		0.773		0.920	
95% CI	0.891-1.000		0.889-0.989		0.567-0.978		0.875-0.965	

**Table 3.** Comparison of clinical data between the two groups of patients [mean  $\pm$  SD/M ( $P_{25}$ ,  $P_{75}$ )]

Patient characteristics	Endometrial carcinoma group (n=140)	Endometrial disease group (n=67)	t/Z	P
Age, year	51.16 $\pm$ 7.61	51.70 $\pm$ 9.00	0.445	0.657
BMI, kg/m <sup>2</sup>	23.92 $\pm$ 2.54	23.36 $\pm$ 2.60	1.475	0.142
CA125, U/mL	33.69 (19.56, 50.21)	17.55 (11.25, 29.12)	5.243	<0.001
CA19-9, U/mL	22.34 (13.67, 33.64)	19.56 (6.80, 28.34)	2.437	0.015
HE4, pmol/L	86.45 $\pm$ 34.80	52.03 $\pm$ 17.95	7.589	<0.001
VEGF, ng/L	322.61 $\pm$ 88.40	227.05 $\pm$ 76.73	7.549	<0.001
D-dimer, ng/mL	383.57 $\pm$ 151.46	284.18 $\pm$ 96.26	4.893	<0.001
RBC, 10 <sup>9</sup> /L	3.74 (3.11, 4.76)	4.26 (3.11, 5.13)	1.158	0.247
MCH, g/L	128.35 $\pm$ 1.72	128.63 $\pm$ 1.78	1.082	0.280
MCV, fL	87.34 $\pm$ 1.42	87.53 $\pm$ 1.06	0.971	0.332
RDW, fL	44.58 $\pm$ 2.89	44.10 $\pm$ 3.16	1.078	0.282
PLT, 10 <sup>9</sup> /L	223.32 $\pm$ 66.94	225.33 $\pm$ 53.89	0.214	0.831
MPV, fL	12.25 (11.00, 13.05)	11.60 (10.70, 12.60)	2.369	0.018
PDW, fL	13.81 $\pm$ 2.13	14.28 $\pm$ 1.96	1.529	0.128
WBC, 10 <sup>9</sup> /L	6.53 $\pm$ 1.51	6.50 $\pm$ 1.22	0.169	0.866
N, 10 <sup>9</sup> /L	4.02 $\pm$ 1.32	3.20 $\pm$ 0.95	4.548	<0.001
L, 10 <sup>9</sup> /L	1.96 (1.54, 2.40)	1.81 (1.49, 2.19)	1.469	0.142
MON, 10 <sup>9</sup> /L	0.33 $\pm$ 0.12	0.33 $\pm$ 0.10	0.315	0.753
NLR	2.04 (1.53, 2.72)	1.75 (1.28, 2.36)	2.578	0.010
PLR	109.15 (83.15, 149.35)	120.63 (94.78, 152.02)	1.342	0.180
MLR	0.17 (0.11, 0.23)	0.17 (0.13, 0.24)	0.619	0.536

Note: BMI: body mass index; VEGF: vascular endothelial growth factor; HE4: human epididymis secretory protein 4; RBC: red blood cell count; MCH: mean corpuscular hemoglobin content; MCV: mean corpuscular volume; RDW: red blood cell distribution width; PLT: platelet count; MPV: mean platelet volume; PDW: platelet distribution width; WBC: white blood cell count.

cacy of HE4 and VEGF were comparable. The diagnostic efficiency of CA19-9 was low, and its AUC value was significantly lower than that of HE4, VEGF and CA125 (Table 5).

### CART model

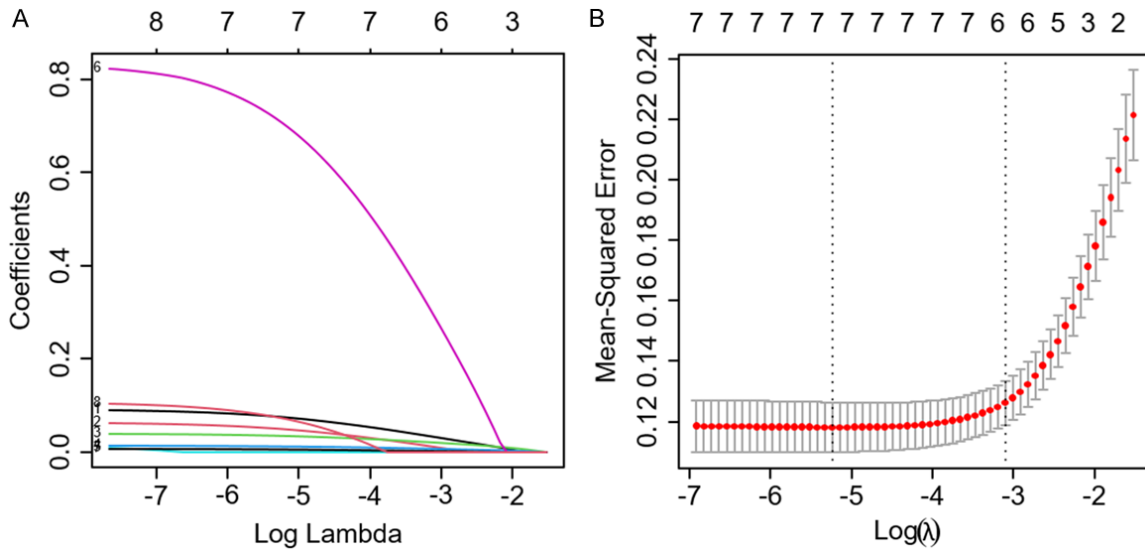
The diagnostic results of endometrial biopsy and 6 diagnostic indicators (CA125, CA19-9, HE4, VEGF, D-dimer, and N) were included in

the CART model. Two effective diagnostic indicators were screened out by the model, which were the diagnostic results of endometrial biopsy and HE4. The model generated three diagnostic rules (Figure 4).

The ROC curve depicted an AUC of 0.949 (95% CI: 0.914-0.985) for the training set and an AUC of 0.942 (95% CI: 0.885-1.000) for the validation set. The model performed well in terms of



## Diagnosis of endometrial carcinoma



**Figure 2.** LASSO regression screening diagnostic indicators. A. The selection path diagram of LASSO regression. B. The change curve of cross-validation under different penalty intensity.

**Table 4.** Multivariate Logistic regression screening diagnostic indicators

Factor	$\beta$	S.E.	Wald $\chi^2$	P	OR	95% CI
CA125	0.092	0.023	15.484	<0.001	1.096	1.047-1.148
CA19-9	0.065	0.023	8.102	0.04	1.067	1.020-1.115
HE4	0.041	0.010	15.700	<0.001	1.041	1.021-1.062
VEGF	0.013	0.003	18.418	<0.001	1.014	1.007-1.020
D-dimer	0.008	0.002	10.129	0.001	1.008	1.003-1.012
MPV	0.107	0.177	0.367	0.545	1.113	0.787-1.574
N	0.833	0.300	7.719	0.005	2.301	1.278-4.142
NLR	0.013	0.338	0.001	0.969	1.013	0.523-1.964

Note: CA125: carbohydrate antigen 125; CA19-9: carbohydrate antigen 19-9; VEGF: vascular endothelial growth factor; HE4: human epididymis secretory protein 4; MPV: mean platelet volume; N: absolute neutrophil count; NLR: neutrophil to lymphocyte ratio.

discriminating (**Figure 5**). The calibration curve showed that the diagnostic outcomes for EC and endometrial lesions in both the training and validation sets of the CART model were consistent with the pathological examination results of patients after hysteroscopic curettage (**Figure 6**).

### *Comparison of diagnostic efficacy between single diagnostic index and CART model*

Among the seven single diagnostic indexes, the diagnostic efficiency of endometrial biopsy via disposable uterine tissue suction tube was the highest, with AUC=0.920, sensitivity =0.914, and specificity =0.925. The diagnostic capa-

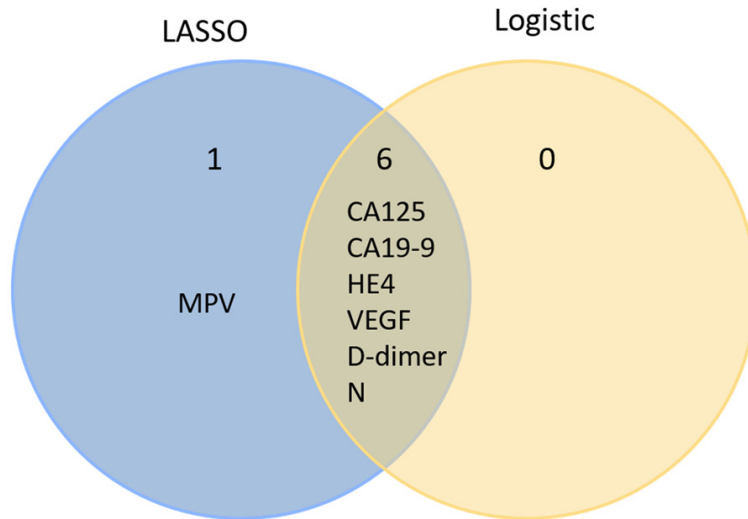
bility of the CART model exceeded that of individual tests, with AUC=0.949, sensitivity = 0.979, and specificity =0.896, demonstrating a strong diagnostic performance (**Table 6**).

### **Discussion**

EC is the most common type of gynecological malignant tumor. Patients with early EC have a 5-year survival rate of more than 85%, while patients with advanced EC have a 5-year survival rate of less than 35% [10, 11]. However, due to the variety of endometrial lesions and the complex structure, the diagnostic efficacy of a single examination is often insufficient, which is easy to cause misdiagnosis and missed diagnosis of EC [8]. Therefore, combined multi-index detection is of great significance for early diagnosis of endometrium, improvement of survival rate and prognosis of patients.

In this study, the disposable uterine tissue suction tube was used for endometrial biopsy. The pathological diagnosis coincidence rate was 90.34% (187/207), showing a high efficiency as a single detection of EC. Cai et al. [4] pointed out that using endometrial biopsy, the diagnosis rate of endometrial atypical

## Diagnosis of endometrial carcinoma



**Figure 3.** Wayne diagram of the common diagnostic indexes in LASSO regression and Logistic regression (CA125: carbohydrate antigen 125; CA19-9: carbohydrate antigen 19-9; VEGF: vascular endothelial growth factor; HE4: human epididymis secretory protein 4; MPV: mean platelet volume; N: absolute neutrophil count).

**Table 5.** Comparison of the efficacy of serological indicators in the diagnosis of endometrial carcinoma

Index	Sensitivity	Specificity	AUC	95% CI
CA125	0.393	1.000	0.725 <sup>a</sup>	0.658-0.793
CA19-9	0.971	0.284	0.605	0.520-0.689
HE4	0.750	0.761	0.797 <sup>a</sup>	0.739-0.856
VEGF	0.821	0.627	0.791 <sup>a</sup>	0.727-0.856
D-dimer	0.550	0.821	0.701 <sup>b</sup>	0.631-0.771
N	0.564	0.776	0.706	0.634-0.777

Note: Compared with the AUC value of the CA19-9, <sup>a</sup>P<0.05. Compared with the AUC value of the D-dimer, <sup>b</sup>P<0.05. CA125: carbohydrate antigen 125; CA19-9: carbohydrate antigen 19-9; VEGF: vascular endothelial growth factor; HE4: human epididymis secretory protein 4; N: absolute neutrophil count.

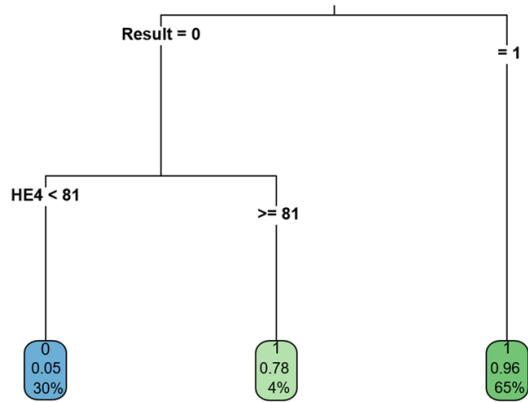
hyperplasia and EC was in good agreement with hysteroscopic curettage. It can be a screening method for EC with great potential. This study found that the sensitivity of endometrial biopsy in the diagnosis of endometrial polyps was low (0.556), and there were many missed diagnoses. Therefore, patients with uterine cavity occupation or highly suspected endometrial polyps suggested by color Doppler ultrasound were not recommended to undergo endometrial biopsy. Relevant study also found that the accuracy of endometrial biopsy with small instruments in the diagnosis of endome-

trial polyps was lower than that of conventional curettage [12]. The single application of endometrial biopsy in the diagnosis of EC has certain limitations. Therefore, combined detection with related serological indicators is particularly important.

In this study, six serological indicators were screened for the diagnosis of EC, which were CA125, CA19-9, HE4, VEGF, D-dimer, and N, respectively. CA125 is essentially an ovarian cancer-related antigen, which is often used to detect epithelial malignant tumors [13]. As a tumor-associated antigen of a class of oligosaccharides, CA19-9 is abundant in the serum of individuals with uterine malignant

tumors [14]. CA125 and CA19-9 are raised to varying degrees in the serum of most patients with EC, but its efficacy as a single diagnosis index is limited for early EC [15, 16]. This study also observed notable elevations in the levels of CA125 and CA19-9 within the EC group. However, relying solely on these two markers for diagnosis yielded unsatisfactory results. The sensitivity of CA125 alone was 0.393, and the specificity of CA19-9 alone was 0.284. HE4 is a whey-acidic protein that is linked to cancer cell development, adhesion, proliferation, and metastasis. HE4 levels have been found to be significantly higher in patients with EC, and it has great sensitivity and specificity in the identification of EC [17, 18]. VEGF is a specific vascular growth factor. As the most effective promoter of vascular endothelial cell division, it is the key to tumorigenesis, invasion and metastasis [19]. Hassan et al. [20] discovered that when endometrial lesions advanced to EC, the expression of VEGF rises considerably. The high expression of VEGF is associated with a higher stage of EC [21]. D-dimer is an important biomarker of fibrinolysis. The invasive growth of tumor tissue causes damage to vascular endothelial cells, which leads to the imbalance between procoagulant and fibrinolytic systems in the human body, and increased level of D-dimer expression in patients' peripheral venous blood [22]. You et al. [23] discovered

## Diagnosis of endometrial carcinoma



**Figure 4.** CART (0: endometrial lesions, 1: endometrial carcinoma). CART: classification and regression tree.

that elevated D-dimer expression was also found in patients with EC. Furthermore, the elevated expression was notably higher in patients with EC than in those with benign endometrial tumors. This study found that the levels of HE4, VEGF, and D-dimer in the EC group were elevated compared to the levels observed in the endometrial lesion group. It is suggested that these three indicators have a certain predictive effect on the development and progression of EC. When HE4 was employed as a solitary diagnostic marker for EC, it yielded an AUC of 0.797, a sensitivity of 0.750, and a specificity of 0.761, indicating a high diagnostic efficiency. Cuesta-Guardiola et al. [24] proposed that serum HE4 levels could serve as an early diagnostic indicator for EC, with its detection efficiency surpassing that of CA125. The study findings revealed that VEGF also exhibited strong diagnostic efficiency as a standalone factor, with an AUC of 0.791, a sensitivity of 0.821, and a specificity of 0.627. The sensitivity of D-dimer was low (0.550). Neutrophils can activate other immune cells and release pro-inflammatory cytokines to promote disease progression [25]. Studies have confirmed that increased N in patients with advanced EC are associated with tumor recurrence and metastasis [26]. According to the findings of this study, the N in the EC group was greatly higher than that in the endometrial lesion group, but its sensitivity for the sole detection of EC was low (0.564), with certain limitations.

Due to the shortcomings of the above individual detection methods for diagnosis, combined detection has become a new direction in clinical

research. Many studies have suggested that combined detection is better than single detection. Combining detection of HE4 with CA125, for example, can aid in the identification of EC [27]. In this study, the results of endometrial biopsy were combined with the above six serological indicators for joint detection, and a CART model was constructed. The model's AUC (0.949), sensitivity (0.979), and specificity (0.896) for diagnosing EC all exhibited enhancements, resulting in a notably elevated diagnostic efficiency compared to the single detection methods. In the validation set, the model also exhibited good diagnostic performance. The CART model screened out two effective diagnostic indicators, results of endometrial biopsy and HE4 level, and generated a total of three diagnostic rules, which were intuitive and easy to explain, thus convenient for extensive promotion in clinical practice.

Certain limitations are present within this study. It is conducted as a single-center retrospective study, potentially introducing some degree of selection bias. At the same time, there is no age stratification of the included patients, so it is impossible to verify whether the model applies to women of different ages. In the future, a multi-center prospective study will be conducted to establish a comparison of various machine algorithms such as random forest and artificial neural network, and to explore their diagnostic efficacy in women of different ages, so as to provide a reference for optimizing the model.

### Conclusion

In summary, this study established a CART model to distinguish EC and endometrial lesions based on the diagnostic results of endometrial biopsy and six serological indicators (CA125, CA19-9, HE4, VEGF, D-dimer, and N), and verified the efficiency of the model, which was found to be an intuitive and simple tool for clinical diagnosis of EC.

### Acknowledgements

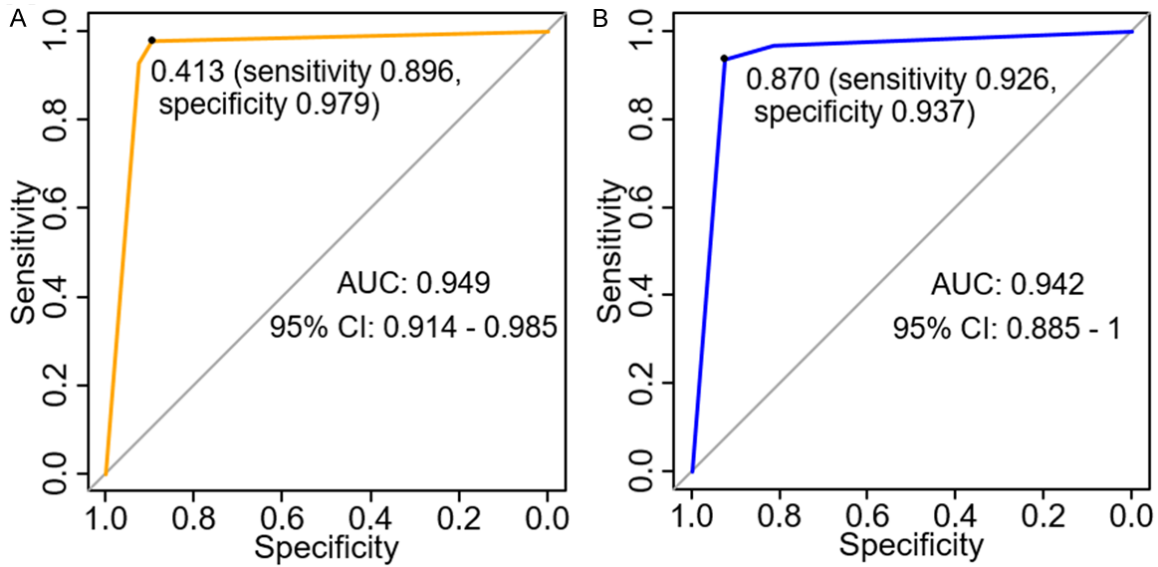
This work was supported by Changde City science and technology innovation project (2022ZD21).

### Disclosure of conflict of interest

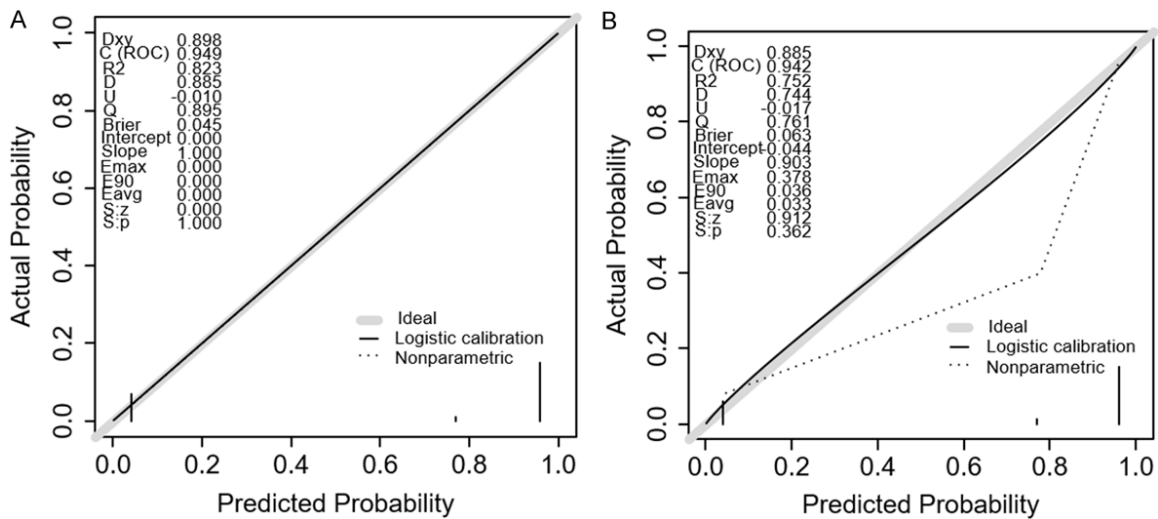
None.



## Diagnosis of endometrial carcinoma



**Figure 5.** ROC curve of CART model. A. Training set. B. Validation set. CART: classification and regression tree.



**Figure 6.** Calibration curve of CART model. A. Training set. B. Validation set. CART: classification and regression tree.

**Table 6.** Comparison of diagnostic efficacy of single diagnostic index and CART model

Index	Sensitivity	Specificity	AUC	95% CI	Z	P
CART model	0.979	0.896	0.949	0.914-0.985	-	-
Endometrial biopsy	0.914	0.925	0.920	0.875-0.965	3.791	<0.001
CA125	0.393	1.000	0.725	0.658-0.793	7.828	<0.001
CA19-9	0.971	0.284	0.605	0.520-0.689	8.963	<0.001
HE4	0.750	0.761	0.797	0.739-0.856	6.536	<0.001
VEGF	0.821	0.627	0.791	0.727-0.856	6.177	<0.001
D-dimer	0.550	0.821	0.701	0.631-0.771	7.991	<0.001
N	0.564	0.776	0.706	0.634-0.777	7.762	<0.001

Note: The *P* value in the table is the AUC value of each index compared with the AUC value of the CART model. CART: classification and regression tree.

## Diagnosis of endometrial carcinoma

**Address correspondence to:** Liyang Wang, Department of Gynecology, Changde Hospital, Xiangya School of Medicine, Central South University, No. 1688 Moon Avenue, Wuling District, Changde 415000, Hunan, China. Tel: +86-0736-7788404; E-mail: wanglycdyy@163.com

### References

- [1] Oaknin A, Bosse TJ, Creutzberg CL, Giorelli G, Harter P, Joly F, Lorusso D, Marth C, Makker V, Mirza MR, Ledermann JA and Colombo N; ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Endometrial cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* 2022; 33: 860-877.
- [2] Huvila J, Pors J, Thompson EF and Gilks CB. Endometrial carcinoma: molecular subtypes, precursors and the role of pathology in early diagnosis. *J Pathol* 2021; 253: 355-365.
- [3] Berger AA, Dao F and Levine DA. Angiogenesis in endometrial carcinoma: therapies and biomarkers, current options, and future perspectives. *Gynecol Oncol* 2021; 160: 844-850.
- [4] Cai L, Liu FM, Ren L and Zhou J. A comparative study between endometrial aspiration biopsy and hysteroscopic curettage. *Acta Academiae Medicinae Xuzhou* 2019; 39: 30-33.
- [5] Cohen SA, Pritchard CC and Jarvik GP. Lynch syndrome: from screening to diagnosis to treatment in the era of modern molecular oncology. *Annu Rev Genomics Hum Genet* 2019; 20: 293-307.
- [6] Lin D, Wang H, Liu L, Zhao L, Chen J, Tian H, Gao L, Wu B, Zhang J, Guo X and Hao Y. IETA ultrasonic features combined with GI-RADS classification system and tumor biomarkers for surveillance of endometrial carcinoma: an innovative study. *Cancers (Basel)* 2022; 14: 5631.
- [7] Wang L, Zhu L, Jiang J, Wang L and Ni W. Decision tree analysis for evaluating disease activity in patients with rheumatoid arthritis. *J Int Med Res* 2021; 49: 3000605211053232.
- [8] Cusimano MC, Vicus D, Pulman K, Maganti M, Bernardini MQ, Bouchard-Fortier G, Laframboise S, May T, Hogen LF, Covens AL, Gien LT, Kupets R, Rouzbahman M, Clarke BA, Mirkovic J, Cesari M, Turashvili G, Zia A, Ene GEV and Ferguson SE. Assessment of sentinel lymph node biopsy vs lymphadenectomy for intermediate- and high-grade endometrial cancer staging. *JAMA Surg* 2021; 156: 157-164.
- [9] Koh WJ, Abu-Rustum NR, Bean S, Bradley K, Campos SM, Cho KR, Chon HS, Chu C, Cohn D, Crispens MA, Damast S, Dorigo O, Eifel PJ, Fisher CM, Frederick P, Gaffney DK, George S, Han E, Higgins S, Huh WK, Lurain JR 3rd, Mariani A, Mutch D, Nagel C, Nekhlyudov L, Fader AN, Remmenga SW, Reynolds RK, Tillmanns T, Ueda S, Wyse E, Yashar CM, McMillian NR and Scavone JL. Uterine neoplasms, version 1.2018, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* 2018; 16: 170-199.
- [10] Morrison J, Balega J, Buckley L, Clamp A, Crosbie E, Drew Y, Durrant L, Forrest J, Fotopoulou C, Gajjar K, Ganesan R, Gupta J, Hughes J, Miles T, Moss E, Nanthakumar M, Newton C, Ryan N, Walther A and Taylor A. British Gynaecological Cancer Society (BGCS) uterine cancer guidelines: recommendations for practice. *Eur J Obstet Gynecol Reprod Biol* 2022; 270: 50-89.
- [11] Zhang S, Gong TT, Liu FH, Jiang YT, Sun H, Ma XX, Zhao YH and Wu QJ. Global, regional, and national burden of endometrial cancer, 1990-2017: results from the global burden of disease study, 2017. *Front Oncol* 2019; 9: 1440.
- [12] Wang T, Jiang R, Yao Y, Wang Y, Liu W, Qian L, Li J, Weimer J and Huang X. Endometrial cytology in diagnosis of endometrial cancer: a systematic review and meta-analysis of diagnostic accuracy. *J Clin Med* 2023; 12: 2358.
- [13] Zhang M, Cheng S, Jin Y, Zhao Y and Wang Y. Roles of CA125 in diagnosis, prediction, and oncogenesis of ovarian cancer. *Biochim Biophys Acta Rev Cancer* 2021; 1875: 188503.
- [14] Bian J, Sun X, Li B and Ming L. Clinical significance of serum HE4, CA125, CA724, and CA19-9 in patients with endometrial cancer. *Technol Cancer Res Treat* 2017; 16: 435-439.
- [15] Ma Y and Shao X. Uterine fibroids with positive 18F-FDG PET/CT image and significantly increased CA19-9: a case report. *Medicine (Baltimore)* 2017; 96: e9421.
- [16] Wan Q, Liu Y, Lv B and Chen X. Correlation of molecular tumor markers CA125, HE4, and CEA with the development and progression of epithelial ovarian cancer. *Iran J Public Health* 2021; 50: 1197-1205.
- [17] Behrouzi R, Barr CE and Crosbie EJ. HE4 as a biomarker for endometrial cancer. *Cancers (Basel)* 2021; 13: 4764.
- [18] Liu J, Han L, Sun Q, Li Y and Niyazi M. Meta-analysis of the diagnostic accuracy of HE4 for endometrial carcinoma. *Eur J Obstet Gynecol Reprod Biol* 2020; 252: 404-411.
- [19] Yu H, Dejizhuoga, Huang W, Wang D, Gamaquzhen, Jia X and Feng H. The expression and clinical significance of sphingosine kinase 1 and vascular endothelial growth factor in endometrial carcinoma. *Emerg Med Int* 2022; 2022: 6716143.
- [20] Hassan WA and Ibrahim R. Expression of CD117, CD34, and VEGF proteins in progression from endometrial hyperplasia to endome-

## Diagnosis of endometrial carcinoma

- trioid carcinoma. *Int J Clin Exp Pathol* 2020; 13: 2115-2122.
- [21] Sunita BS, Sen A and Suhag V. To evaluate immunoreactivity of cyclooxygenase-2 in cases of endometrial carcinoma and correlate it with expression of p53 and vascular endothelial growth factor. *J Cancer Res Ther* 2018; 14: 1366-1372.
- [22] Carugno J. Clinical management of vaginal bleeding in postmenopausal women. *Climacteric* 2020; 23: 343-349.
- [23] You SJ, Zhang HL, Rao JH, Zhou AX, Dong ZZ, Chen GL, Lv YC and Su ML. The clinical value of preoperative NLR, CA125 and plasma D-dimer in the diagnosis of endometrial carcinoma. *China Health Standard Management* 2022; 13: 76-79.
- [24] Cuesta-Guardiola T, Carretero AQ, Martinez-Martinez J, Cunarro-Lopez Y, Pereira-Sanchez A, Fernandez-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.
- [25] Ural UM, Sehitoglu I, Tekin YB and Sahin FK. Neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios in patients with endometrial hyperplasia and endometrial cancer. *J Obstet Gynaecol Res* 2015; 41: 445-448.
- [26] Muzykiewicz KP, Iwanska E, Janeczek M, Glanowska I, Karolewski K and Blecharz P. The analysis of the prognostic value of the neutrophil/lymphocyte ratio and the platelet/lymphocyte ratio among advanced endometrial cancer patients. *Ginekol Pol* 2021; 92: 16-23.
- [27] Wu Q, Bai SN, Song LY, Wu WF and Han LN. Diagnostic value of serum human epididymis protein 4, carbohydrate antigen 125 and their combination in endometrial cancer: a meta-analysis. *Medicine (Baltimore)* 2023; 102: e34737.