

# Development and External Validation of Machine Learning Model to Predict Live Birth Following Assisted Reproductive Technology in Women with Ovarian Endometriomas: A Decision-Support Tool

Yifei Sun<sup>1-7,\*</sup>, Zijing Wang<sup>1-7,\*</sup>, Jiayi Zhou<sup>1-7</sup>, Linlin Cui<sup>1-8</sup>, Huidan Wang<sup>1-7</sup>

<sup>1</sup>State Key Laboratory of Reproductive Medicine and Offspring Health, Center for Reproductive Medicine, Institute of Women, Children and Reproductive Health, Shandong University, Jinan, Shandong, 250012, People's Republic of China; <sup>2</sup>National Research Center for Assisted Reproductive Technology and Reproductive Genetics, Shandong University, Jinan, Shandong, 250012, People's Republic of China; <sup>3</sup>Key Laboratory of Reproductive Endocrinology, Shandong University, Ministry of Education, Jinan, Shandong, 250012, People's Republic of China; <sup>4</sup>Shandong Technology Innovation Center for Reproductive Health, Jinan, Shandong, 250012, People's Republic of China; <sup>5</sup>Shandong Provincial Clinical Research Center for Reproductive Health, Jinan, Shandong, 250012, People's Republic of China; <sup>6</sup>Shandong Key Laboratory of Reproductive Research and Birth Defect Prevention, Jinan, Shandong, 250012, People's Republic of China; <sup>7</sup>Research Unit of Gametogenesis and Health of ART-Offspring, Chinese Academy of Medical Sciences (No. 2021RU001), Jinan, Shandong, 250012, People's Republic of China; <sup>8</sup>Center for Reproductive Medicine, The Second Qilu Hospital of Shandong University, Jinan, Shandong, 250012, People's Republic of China

\*These authors contributed equally to this work

Correspondence: Huidan Wang, Email wanghd1984@126.com

**Background:** Ovarian endometriomas damage the ovarian structure, alter ovarian inflammation, and impair ovarian reserve. Given the conflicting results, determining an optimal reproductive strategy for women with endometriomas—whether expectant management, medication, surgery, or assisted reproductive technology (ART)—remains challenging.

**Objective:** This study aims to preliminarily develop and validate clinically applicable decision-support tools by training, testing, and validating an automated machine learning (ML) model to predict the likelihood of live birth following ART in women with endometriomas.

**Methods:** The derivation and testing cohort included 1705 women, and the external validation cohort included 1475 women with ovarian endometriomas following ART retrospectively. Two ML models were developed and validated to predict the probability of live birth. Model performance was evaluated using the area under the curve (AUC), accuracy, sensitivity, specificity, F1 score, and Brier score. The SHapley Additive exPlanations (SHAP) method was employed to interpret feature importance.

**Results:** Comparing seven ML algorithms, the Extreme Gradient Boosting (XGBoost) demonstrated superior predictive performance both in model-1 and model-2, achieving an AUC of 0.90 [95% confidence interval (CI): 0.88–0.92] and 0.88 (95% CI: 0.86–0.89) in test-datasets and 0.80 (95% CI: 0.76–0.83) and 0.69 (95% CI: 0.65–0.73) in external validation cohort. The SHAP analysis revealed that the age and features associated with ovarian reserve had strong predictive power and the ovarian endometriomas had limited predictive power.

**Conclusion:** Model-2, which uses only pre-ART variables, can support reproductive strategy selection prior to ART initiation. Conversely, Model-1 is designed to support embryo transfer strategy option after oocyte retrieval, incorporating post-ART data. Although both models show promise as decision-support tools for personalizing infertility treatment in women with endometriomas, their clinical implementation awaits confirmation from prospective, multicenter validation.

**Keywords:** machine learning model, assisted reproductive technology, ovarian endometriomas, live birth, SHAP interpretation

## Introduction

Endometriosis (EM) is a common gynecological disorder characterized by the presence of endometrial-like tissue outside the uterine cavity, often associated with chronic pelvic pain and infertility.<sup>1,2</sup> Ovarian endometriomas affect 17–44% of



women with endometriosis.<sup>3</sup> Unlike peritoneal lesions, these cysts directly compromise ovarian reserve through local inflammation, oxidative stress, and follicle loss from cyst expansion.<sup>4,5</sup> For women with endometriomas seeking conception, the choice between direct recourse to assisted reproductive technology (ART) or prior surgical intervention is uncertain.<sup>6,7</sup> Surgery may relieve symptoms but risks further ovarian damage,<sup>8–11</sup> while immediate ART must account for potentially diminished response.<sup>12–14</sup>

Existing prediction tools offer limited guidance for this population. The Endometriosis Fertility Index (EFI) predicts natural conception after surgery but requires operative findings, excluding non-surgical candidates.<sup>15–17</sup> Yet the statistical tools commonly used in ART research—most notably logistic regression—operate under a linearity assumption that sits uneasily with the intricate, non-linear interplay of factors shaping treatment outcomes.<sup>18</sup>

Machine learning (ML) can capture these complex patterns without pre-specified assumptions,<sup>19</sup> yet few studies have developed validated models specifically for women with endometriomas.<sup>20–22</sup> Most existing models predict pregnancy per transfer<sup>23,24</sup> rather than cumulative live birth (CLB) per retrieval, the outcome patients prioritize,<sup>25,26</sup> thereby limiting their generalizability and readiness for clinical use.

We developed and externally validated two machine learning models tailored to this population. Model-1 incorporates post-ART laboratory parameters to support embryo transfer decisions. In contrast, Model-2 uses only pre-ART clinical variables to inform the spontaneous conception-versus-ART choice after endometriomas surgery. By comparing multiple algorithms and applying SHAP for interpretability, we aimed to produce transparent tools that provide personalized probability estimates and support shared decision-making along the entire treatment pathway.

## Materials and Methods

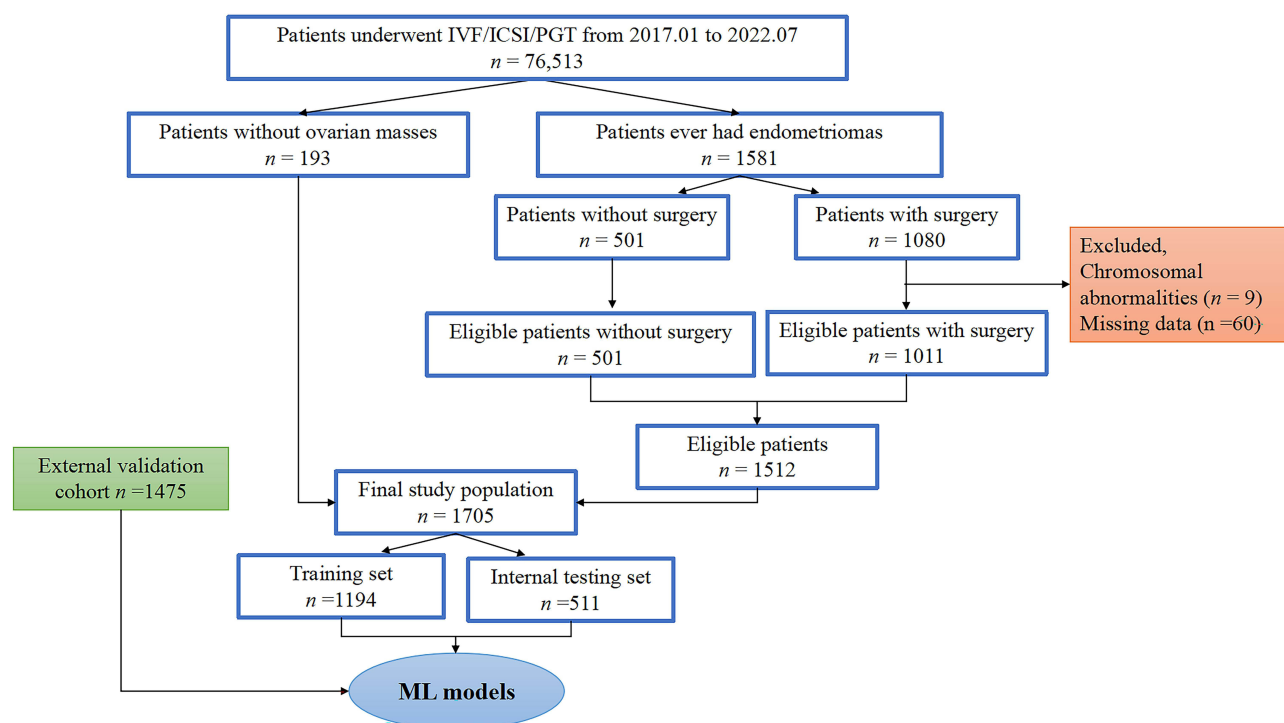
### Study Population

The study population was retrospectively recruited from January 2017 to July 2022 at the Hospital for Reproductive Medicine Affiliated to Shandong University. Baseline characteristics of the women were retrieved from the medical record system. A total of 1581 patients ever had ovarian endometriomas (1080 patients with surgery and 501 patients without surgery) and 193 women without ovarian masses who underwent in vitro fertilization (IVF)/intracytoplasmic sperm injection (ICSI)/preimplantation genetic testing (PGT) were recruited. The males or females with chromosomal abnormalities, or for whom key features associated with live birth had a high amount of missing data (>30%) were excluded. Ultimately, 1705 eligible women were included in the study dataset, which divided into 70% train (n=1194) and 30% test (n=511) datasets. We also recruited 217 patients with a history of ovarian endometriomas and 1331 women without ovarian masses from March 2022 to December 2023 at the Center for Reproductive Medicine, the Second Hospital, Cheeloo College of Medicine, Shandong University. With the same inclusion and exclusion criteria, 1475 women were preserved as an external validation dataset. The study flowchart was shown in [Figure 1](#) and the specific flowchart for the selection of external validation patients was shown in [Figure S1](#).

### Exposure

The main exposure measure was ovarian endometriotic cysts. Bilateral, compared to unilateral, may result in a greater negative effect on ovarian reserve both before and after ovarian cystectomy for endometriomas in some researches.<sup>27,28</sup> A review on treatments of endometriomas showed that each approach has varying effects on ovarian reserve, spontaneous pregnancy rates and recurrence rates.<sup>29</sup> The surgical therapy of endometriomas was achieved via sclerotherapy, acupuncture, ablation, cystectomy and oophorectomy in the women of our study. Therefore, in order to predict the impact of the above factors on the fertility of women with endometriomas, we classified seven endometrioma categories according to the surgical approaches and unilateral/bilateral endometriomas: 1) women without ovarian masses, 2) unilateral endometriomas, 3) bilateral endometriomas, 4) sclerotherapy and acupuncture, 5) unilateral cystectomy and ablation (electrosurgical/laser), 6) bilateral cystectomy and ablation (electrosurgical/laser), and 7) oophorectomy.

Body mass index (BMI) was computed using the ratio of weight to the square of height. The basic endocrinological profiles [follicle stimulating hormone (FSH), luteinizing hormone (LH), estradiol (E2) and anti-Müllerian hormone (AMH)] and antral follicle count (AFC) were tested during the first three days of the menstrual cycle. FSH, LH and E2



**Figure 1** The flowchart for the selection of the study population.

**Abbreviations:** IVF, in vitro fertilization; ICSI, intracytoplasmic sperm injection; PGT, preimplantation genetic testing; ML, machine learning.

were tested by chemiluminescence immunoassays (Roche Diagnostics, Germany) with intra- and inter-assay coefficients of variation of less than 10%. AMH was tested using an enzyme-linked immunosorbent assay (Ansh Labs, Webster, USA). All blood samples were stored at  $-80^{\circ}\text{C}$  until detection. The AFC was defined as the count of follicles with a diameter of 2–9 mm in both ovaries. The stimulation protocols included natural cycle or modified natural cycle, gonadotrophin releasing hormone (GnRH) agonist protocols, GnRH antagonist protocol and other protocols using recombinant FSH, urine-derived human menopausal gonadotrophin (HMG), or GnRH antagonist.

## Study Outcome

The aim of our study was to assess automated ML models as assisted tools to enhance individualized clinical strategies by predicting probability of live birth following ART among patients with endometriomas. The primary outcome of our study was whether the patients obtained a live birth or more in a single ART cycle. The CLB was defined as whether women had a live birth at 28 weeks or more of gestation after transfers of all embryos or up to three blastocysts in the first ART cycle within 1 year.<sup>25</sup>

## Data Processing and Feature Selection

In this research, we selected features based on the availability, clinical experience and relevant frontier literature. Finally, a total of 20 variables were included in our study combined with the inclusion and exclusion criteria. These 20 variables were integral in facilitating the feature processing and construct ML models. Our analysis primarily focused on CLB, categorizing them into successful-live-birth or non-live-birth. In addition, we also examined demographic characteristics (such as women age, BMI, education level and occupation), ultrasound (AFC), basic endocrinology (FSH, LH, E2 and AMH), surgery (treatment of endometriomas and hysteroscopy) and ART relevant features [such as length of infertility, ovarian stimulation, days of stimulation, Gn starting dose, total Gn dose, endometrial thickness on trigger day, number of retrieved oocytes, 2-pronuclea (2PN) zygotes and blastocysts]. All predict features were measured long before the outcome (CLB). No post-outcome information was included in predictors.

In this study, the overall rate of missing data was low: 0.27% in model-1 (23 missing values out of 1705×5 variables) and 0.03% in model-2 (5 missing values out of 1705×10 variables). The specific distributions of missing data values in model-1 and model-2 were showed in [Figure S2](#). We employed available case analysis (pairwise deletion) to handle missing values. This approach maximizes statistical power while maintaining transparency about data usage. Given the minimal proportion of missing data, this simple approach avoids the potential complexities and assumptions associated with imputation methods while yielding reliable estimates. We employed the least absolute shrinkage and selection operator (LASSO) regression to select the most relevant features and discard extraneous variables<sup>30</sup> to prevent overfitting.<sup>31</sup>

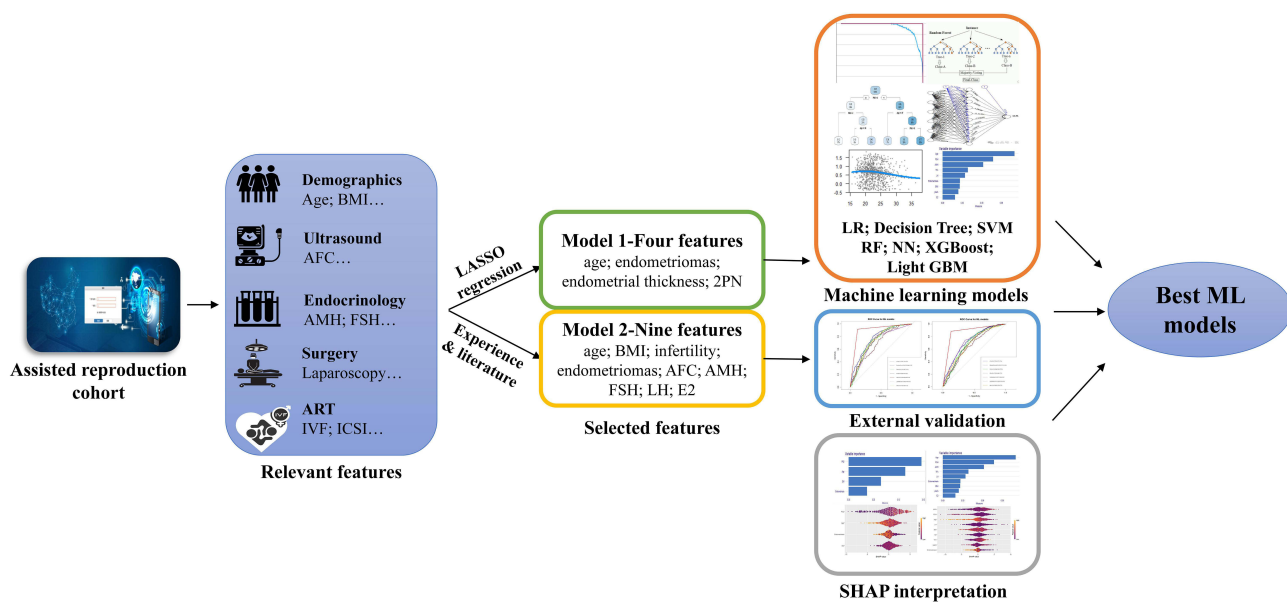
## Construction, Evaluation and Interpretation of ML Models

In this study, we randomly partitioned the study dataset into two groups using a 70% to 30% train-test split. The primary purpose of the training dataset was to develop ML models, while the testing dataset was reserved exclusively for assessing the predictive capabilities of the ML models. We used seven prevalent algorithms to develop ML models to predict CLB of a single ART cycle in patients with endometriomas: Logistic Regression (LR), Decision Tree, Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM) and Neural Network (NN). Five cross-validations were conducted in order to determine the optimal hyperparameters for seven ML models. The training dataset is preprocessed applying a mix of random under-sampling and Synthetic Minority Over-sampling Technique (SMOTE) to address the issue of positive and negative sample imbalance.<sup>32</sup> The SMOTE was applied to the training dataset to mitigate potential bias toward the majority class and to enhance the robustness of minority-class recognition.<sup>33</sup> The testing dataset was utilized to evaluate the models predictable when the model selection and training procedures completed.

Model performance was assessed using AUC, accuracy, sensitivity, specificity, F1 score, and Brier score. The Brier score is an evaluation metric that quantifies the discrepancy between predicted probabilities and actual outcomes. A lower value indicates superior model performance.<sup>34</sup>

The contribution of each variable to the prediction results of ML models was determined using the SHAP method by drawing a swarm diagram. SHAP evaluations of selected cases showed how much a feature affected a particular sample and helped us understand the decision -making process of ML models. The overall workflow of the study is presented in [Figure 2](#).

All statistical analyses were performed using the R software (version: 4.2.2).



**Figure 2** The overall flowchart of the study.

**Abbreviations:** BMI, body mass index; AFC, antral follicle count; AMH, anti-Müllerian hormone; FSH, follicle-stimulating hormone; LH, luteinizing hormone; E2, estradiol; LASSO, Least absolute shrinkage and selection operator; 2PN, 2-pronuclear zygotes; LR, Logistic Regression; SVM, Support Vector Machine; RF, Random Forest; NN, Neural Network; XGBoost, Extreme Gradient Boosting; Light GBM, Light Gradient Boosting Machine; SHAP, SHapley Additive exPlanations.

## Results

### Sample Baseline Characteristics

A total of 1581 patients with endometriosis cysts were included according to the inclusion and exclusion criteria. Subsequently, 9 patients with abnormal chromosomes and 60 patients with missing data were omitted. We also included 193 patients without any ovarian cysts or ovarian tumor as control. Finally, 1705 patients were included as delineated in [Figure 1](#) and 1475 patients were included as external validation in [Figure S1](#), with details presented in [Tables 1](#) and [2](#). Differences in baseline characteristics between derivation and testing cohort and external validation cohort patients were shown in [Table S1](#).

[Table 1](#) showed baseline characteristics and infertility assessment parameters of all patients in derivation and testing database. There were 810 patients (47.5%) obtaining successful live birth in an ART cycle. Notably, younger patients and those who had better ovarian reserve [more AFC, lower FSH and higher AMH levels] were more prone to obtaining successful live births in single ART cycle. Endometriomas effected the outcomes of ART cycle, and the live birth rate of an ART cycle is much higher in patients without endometriomas than patients with endometriomas [64.77% (125/193) vs. 45.30% (685/1512)]. [Table 2](#) showed the comparisons of ovarian stimulation protocols and laboratory outcomes between the two groups. The ovarian stimulation protocol and laboratory signs assessments revealed that the patients

**Table 1** Comparisons of Baseline Characteristics Between the Non-Live-Birth and Successful-Live-Birth Groups

Variables	Patients with Endometriosis Cysts Following ART		
	Total (n=1705)	Non-Live Birth (n=895)	Successful Live Birth (n=810)
Age, years, M (Q1, Q3)	32 (30, 35)	33 (30, 37)	31 (29, 34)
BMI, kg/m <sup>2</sup> , M (Q1, Q3)	22.44 (20.53, 24.88)	22.61 (20.56, 25.22)	22.20 (20.51, 24.45)
Higher education level, n (%)	1074 (62.99)	563 (62.9)	511 (63.1)
Occupation, n (%)			
Unskilled	773 (45.34)	393 (43.9)	380 (46.9)
Semi-skilled	52 (3.05)	28 (3.1)	24 (3.0)
Skilled	880 (51.61)	474 (53.0)	406 (50.1)
AFC, M (Q1, Q3)	10 (6, 14)	8 (5, 12)	11 (8, 16)
Basal FSH (IU/L), M (Q1, Q3)	7.03 (5.89, 8.67)	7.34 (6.14, 9.68)	6.74 (5.73, 7.89)
Basal LH (IU/L), M (Q1, Q3)	5.51 (3.60, 6.36)	4.81 (3.50, 6.39)	4.98 (3.74, 6.34)
Basal E <sub>2</sub> (pg/mL), M (Q1, Q3)	40.50 (29.83, 53.70)	41.10 (29.54, 55.92)	39.85 (30.20, 51.19)
AMH (ng/mL), M (Q1, Q3)	2.12 (1.07, 3.75)	1.51 (0.74, 2.92)	2.80 (1.64, 4.45)
Endometriomas, n (%)			
Non-ovarian masses	193 (11.32)	68 (7.60)	125 (15.4)
Intact unilateral endometriomas	351 (20.59)	180 (20.1)	171 (21.1)
Intact bilateral endometriomas	150 (8.80)	71 (7.9)	79 (9.8)
Sclerotherapy and acupuncture	14 (0.82)	7 (0.8)	7 (0.9)
Unilateral cystectomy and ablation	665 (39.00)	348 (38.9)	317 (39.1)
Bilateral cystectomy and ablation	325 (19.06)	216 (24.1)	109 (13.5)
Oophorectomy	7 (0.41)	5 (0.6)	2 (0.2)
Length of infertility (years), M (Q1, Q3)	3 (2, 4.5)	3.00 (2.00, 5.00)	3.00 (2.00, 4.50)
Hysteroscopy, n (%)			
Normal	1121 (65.75)	566 (65.4)	555 (70.7)
Intrauterine adhesions	50 (2.93)	32 (3.6)	18 (2.22)
Endometrial polyps/submucosal fibroids	460 (26.98)	254 (28.4)	206 (25.4)
Uterine malformations	19 (1.11)	10 (1.1)	9 (1.1)
Scarring diverticulum	12 (0.70)	9 (1.0)	3 (0.4)

**Notes:** M: Median, Q<sub>1</sub>: 1<sup>st</sup> Quartile, Q<sub>3</sub>: 3<sup>rd</sup> Quartile.

**Abbreviations:** ART, assisted reproductive technology; BMI, body mass index; AFC, antral follicle count; FSH, follicle-stimulating hormone; LH, luteinizing hormone; E<sub>2</sub>, estradiol; AMH, anti-Müllerian hormone.

**Table 2** Comparisons of ART Treatment Protocols and Laboratory Outcomes Between the Non-Live-Birth and Successful-Live-Birth Groups

Variables	Patients with Endometriosis Cysts Following ART		
	Total (n=1705)	Non-Live Birth (n=895)	Successful Live Birth (n=810)
Ovarian stimulation protocol, n (%)			
Natural cycle or modified natural cycle	51 (2.99)	49 (5.5)	2 (0.2)
GnRH agonist protocol	1178 (69.09)	561 (62.7)	617 (76.2)
GnRH antagonist protocol	363 (21.29)	192 (21.5)	171 (21.1)
Others	113 (6.63)	93 (10.4)	20 (2.5)
Days of stimulation, M (Q1, Q3)	10 (9, 11)	10 (8, 11)	10 (9, 11)
Gn starting dose (IU), M (Q1, Q3)	175 (150, 225)	200 (150, 225)	150 (150, 225)
Total Gn dose (IU), M (Q1, Q3)	2025 (1500, 2850)	2100 (1500, 3000)	2025 (1500, 2700)
Endometrial thickness on trigger day (cm), M (Q1, Q3)	1.10 (0.95, 1.25)	1.00 (0.90, 1.20)	1.10 (1.00, 1.25)
Number of retrieved oocytes, M (Q1, Q3)	8 (4, 12)	5 (2, 9)	10 (7, 14)
Number of 2PN zygotes, M (Q1, Q3)	5 (2, 8)	3 (1, 6)	6.0 (4, 9)
Number of blastocysts, M (Q1, Q3)	2 (1, 4)	1 (0, 3)	3 (1, 5)

**Notes:** M: Median, Q<sub>1</sub>: 1<sup>st</sup> Quartile, Q<sub>3</sub>: 3<sup>rd</sup> Quartile.

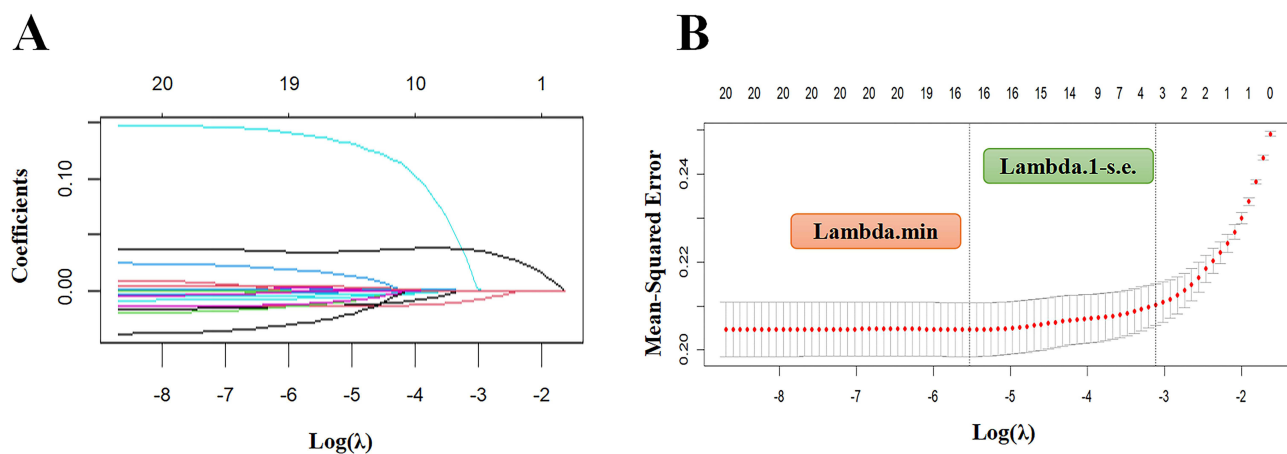
**Abbreviations:** ART, assisted reproductive technology; GnRH, gonadotrophin releasing hormone; 2PN, 2-pronuclear zygotes.

obtaining successful live birth were more likely to use a gonadotrophin releasing hormone (GnRH) agonist protocol with more days of stimulation, thicker endometrium and more retrieved oocytes/2PN zygotes/blastocysts. Conversely, they were demonstrated to use a lower gonadotropin starting dose [150 (150, 225) vs. 200 (150, 225)].

## Feature Selected in ML Models

The feature engineering was conducted to prevent the risk of over-fitting and to screen the important features which impact the ovarian reserve and cumulative live birth in the ART cycle. In order to narrow the candidate features, the least absolute shrinkage and LASSO regression were performed to screen the relevant features of the training database and characteristics of the variable coefficients were displayed in Figure 3. The iterative analysis was performed using a tenfold cross-validation method. LASSO regression effectively reduces the loss function (binomial deviation) by adjusting the regularization coefficient, lambda ( $\lambda$ ) and ultimately yielding a zero coefficient for certain variables.

Of the 20 variables, four were identified as the best predictors and were included in the ML model-1 were patients age, treatment of endometriomas, endometrial thickness and the number of 2PN zygotes (Table S2). These were identified at



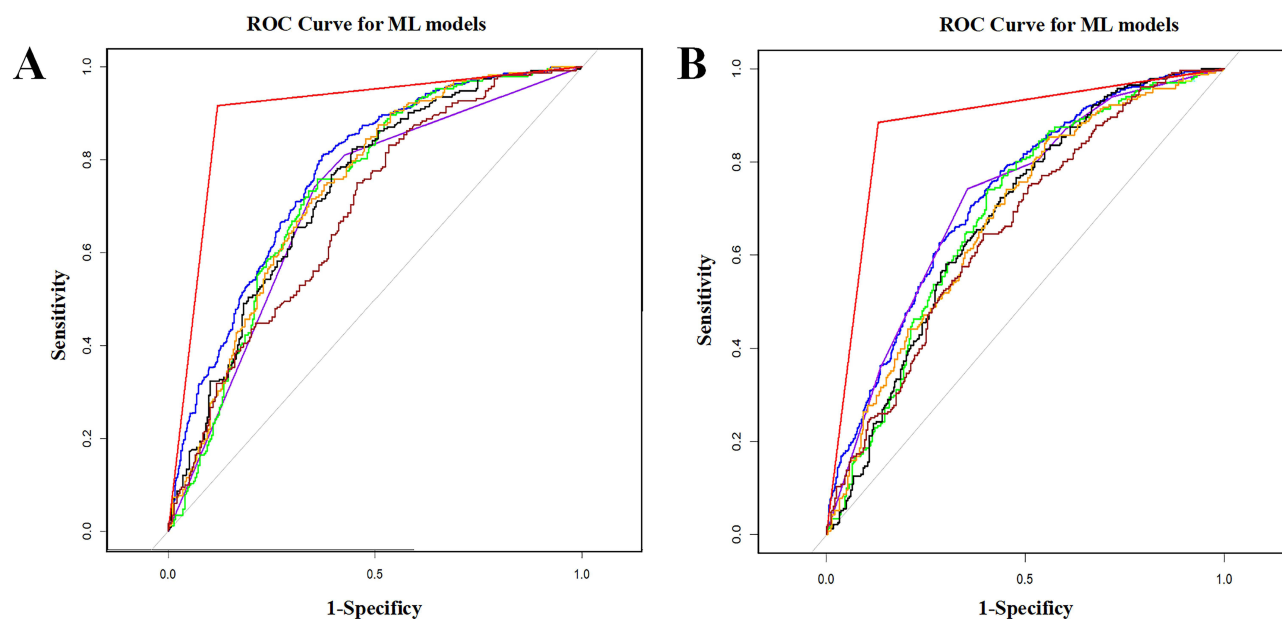
**Figure 3** Lasso regression-based variable screening. (A) Variation characteristics of variable coefficients. (B) The process of selecting the optimal value of the parameter  $\lambda$  in the lasso regression model is carried out by the cross-validation method.

a shrinkage parameter ( $\lambda$ , 1-s. e) of 0.04. In addition, both patients and doctors need to assess the likelihood of obtaining a live birth prior to the ART cycle. Referencing to the features identified at a shrinkage parameter ( $\lambda$ , min) of 0.004, we also selected nine easily available variables which excluded ART parameters (patients age, BMI, duration of infertility, treatment of endometriosis cysts, AFC, AMH, FSH, LH and E2) to constructed ML model-2 (Table S3).

## Comparison of Live Birth Prediction Model Performance in Seven Algorithms

Seven prevalent different ML algorithms – LR, Decision Tree, SVM, RF, XGBoost, LightGBM and NN – were conducted to predict the possibility of obtaining successful live birth in the ART cycles among patients with ovarian endometriomas. The seven ML models were trained employing the training datasets after tuning the hyperparameters, subsequently, their performance was predicted using the testing datasets. To forecast the possibility of CLB in single ART cycle, receiver operating characteristic (ROC) curves were generated from the seven ML models (Figure 4A and B).

Figure 4A displayed the discriminative performance of model-1 trained by the seven algorithms in terms of ROC curves. All seven algorithms showed considerable prediction performance for the possibility of live birth and the XGBoost model exhibiting superior performance with an AUC of 0.90 (95% CI: 0.88–0.92). The remaining models, while still demonstrating good predictive power, ranked as follows in descending order of performance: LR (AUC=0.77, 95% CI: 0.74–0.80), NN (AUC=0.74, 95% CI: 0.69–0.78), SVM (AUC=0.73, 95% CI: 0.69–0.77), Decision Tree (AUC=0.71, 95% CI: 0.67–0.75), RF (AUC=0.71, 95% CI: 0.66–0.75), Light GBM (AUC=0.69, 95% CI: 0.64–0.73) (Table 3 and Figure 4A). To fully assess the model's performance, the detailed performance metrics for the seven algorithms: the sensitivity, specificity, accuracy, recall rate, F1 and Brier score were indicated in Table 3. The XGBoost model exhibited best overall performance (sensitivity: 0.92, specificity: 0.88) and achieved the highest F1 score (0.90) and accuracy (0.90, 95% CI: 0.88, 0.91). Meanwhile, the XGBoost model also boasted the highest recall rate (0.92) and the lowest Brier score (0.14) among all seven algorithms evaluated.



**Figure 4** ROC curves for the ML models. **(A)** ROC curves of the ML models for CLB prediction based on Model-1 with 4 variables. The blue line represents LR, AUC=0.77 (0.74–0.80); the purple line represents Decision Tree, AUC=0.71 (0.67–0.75); the green line represents SVM, AUC=0.73 (0.69–0.78); the black line represents RF, AUC=0.71 (0.66–0.75); the red line represents XGBoost, AUC=0.90 (0.88–0.92); the brown line represents lightGBM, AUC=0.68 (0.64–0.73); the Orange line represents NN, AUC=0.74 (0.69–0.78). **(B)** ROC curves of the ML models for CLB prediction based on Model-2 with 9 variables. The blue line represents LR, AUC=0.73 (0.70–0.75); the purple line represents Decision Tree, AUC=0.66 (0.61–0.70); the green line represents SVM, AUC=0.69 (0.64–0.74); the black line represents RF, AUC=0.73 (0.69–0.78); the red line represents XGBoost, AUC=0.88 (0.86–0.89); the brown line represents lightGBM, AUC=0.66 (0.61–0.71); the orange line represents NN, AUC=0.68 (0.64–0.73).

**Abbreviations:** ROC, receiver operating characteristic; ML, machine learning; CLB, cumulate live birth; AUC, area under the curves; LR, logistic regress; SVM, support vector machine; RF, random forest; XGBoost, extreme gradient boosting; lightGBM, light gradient boosting machine; NN, neural networks.

**Table 3** Performance of the ML Model-1 for Predicting CLB

Model-1	AUC (95% CI)	Sensitivity	Specificity	F1 Score	Accuracy	Recall	Brier
LR	0.77 (0.74–0.80)	0.64	0.74	0.67	0.69 (0.66–0.72)	0.64	0.20
Decision tree	0.71 (0.67–0.75)	0.74	0.65	0.69	0.69 (0.65–0.73)	0.74	0.46
SVM	0.73 (0.69–0.77)	0.74	0.64	0.68	0.68 (0.64–0.72)	0.74	0.47
RF	0.71 (0.66–0.75)	0.73	0.61	0.67	0.67 (0.62–0.71)	0.73	0.50
XGBoost	0.90 (0.88–0.92)	0.92	0.88	0.90	0.90 (0.88–0.91)	0.92	0.14
Light GBM	0.69 (0.64–0.73)	0.64	0.60	0.60	0.62 (0.58–0.66)	0.64	0.49
NN	0.74 (0.69–0.78)	0.72	0.64	0.67	0.67 (0.63–0.71)	0.72	0.46

**Abbreviations:** ML, machine learning; CLB, cumulate live birth; AUC, area under the curves; LR, logistic regress; SVM, support vector machine; RF, random forest; XGBoost, extreme gradient boosting; lightGBM, light gradient boosting machine; NN, neural networks.

The discriminative performance of model-2 trained by the seven algorithms in terms of ROC curves was presented in Figure 4B. The XGBoost model again outperformed the others, with an AUC of 0.88 (95% CI: 0.86–0.89) (Table 4 and Figure 4B). In descending order of performance, the AUCs of RF, LR, SVM, NN, Light GBM and Decision Tree models were 0.73 (95% CI: 0.69–0.78), 0.73 (95% CI: 0.70–0.75), 0.69 (95% CI: 0.64–0.73), 0.68 (95% CI: 0.64–0.73), 0.66 (95% CI: 0.61–0.71), 0.66 (95% CI: 0.61–0.70), respectively (Table 4 and Figure 4B). The XGBoost also exhibited superior overall performance and its sensitivity, specificity, accuracy, recall, F1 and Brier score were 0.89, 0.86, 0.88 (95% CI: 0.86, 0.89), 0.89, 0.88 and 0.15, respectively (Table 4). Although, the prediction effect of model-2 on CLB was lower than that of model-1, the XGBoost algorithm still had good comprehensive performance.

In clinical prediction models, an AUC above 0.8 is generally considered excellent discrimination. Our XGBoost-based Model-1 achieved an outstanding AUC of 0.90, indicating excellent discrimination—substantially higher than the traditional logistic regression (AUC=0.77). The performance gap suggests that the advantage of non-linear ensemble methods for capturing complex relationships in this dataset. Model-2, relying solely on pre-ART variables, also performed well (AUC=0.88), making it suitable for early-cycle counseling.

## External Validation

A critical strength of this study lies in the external validation of our models using an independent cohort from a separate clinical center. This step tests the generalizability of the models beyond the data on which they were trained. Despite inherent differences in baseline characteristics between the derivation and external validation cohorts (Table S1), Model-1 maintained an AUC of 0.80 (95% CI: 0.76–0.83) and Model-2 achieved an AUC of 0.69 (95% CI: 0.65–0.73) (Figure S3). As expected, Model-1 showed better performance retention given its inclusion of cycle-specific ART laboratory parameters. This external validation supports that the models are not overfitted to the training data and may generalize to new populations—an essential criterion for clinical applicability.

**Table 4** Performance of the ML Model-2 for Predicting CLB

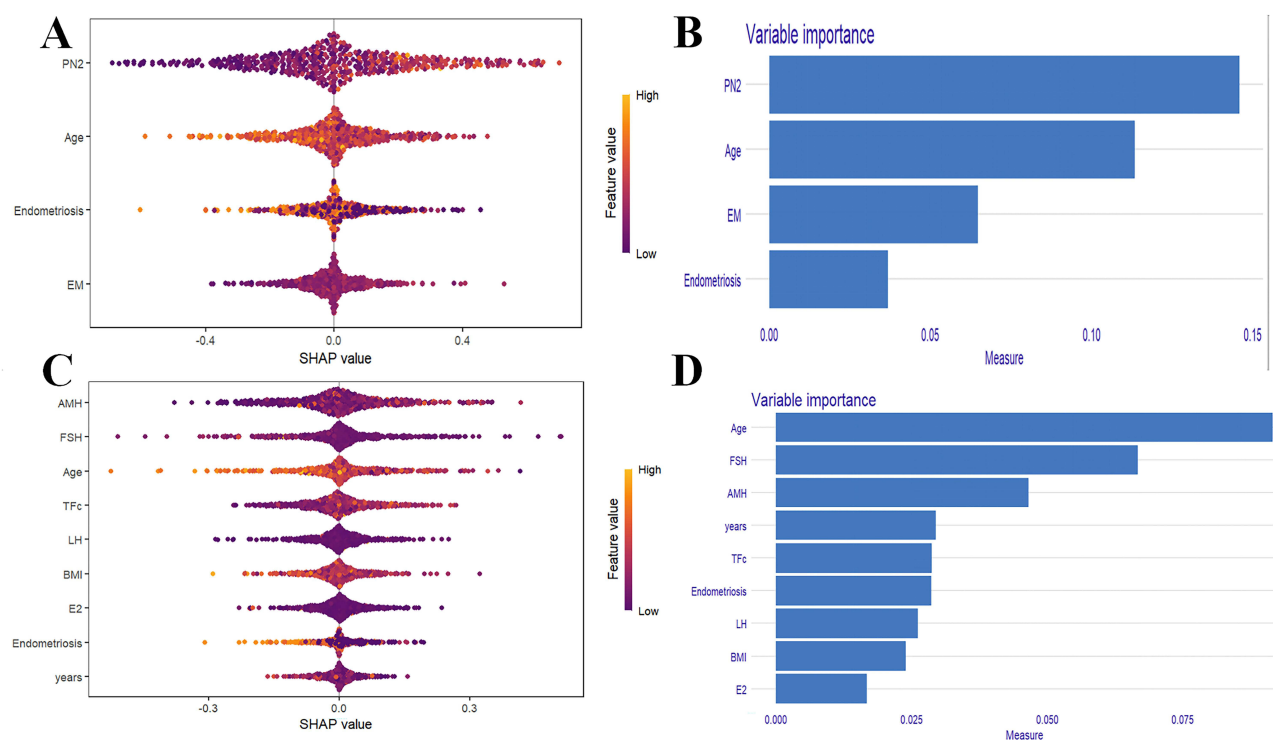
Model-2	AUC (95% CI)	Sensitivity	Specificity	F1 Score	Accuracy	Recall	Brier
LR	0.73 (0.70–0.75)	0.66	0.68	0.66	0.67 (0.64–0.70)	0.66	0.33
Decision tree	0.66 (0.61–0.70)	0.61	0.62	0.59	0.62 (0.58–0.66)	0.61	0.44
SVM	0.69 (0.64–0.73)	0.63	0.65	0.62	0.64 (0.60–0.68)	0.63	0.40
RF	0.73 (0.69–0.78)	0.65	0.62	0.62	0.64 (0.59–0.68)	0.65	0.46
XGBoost	0.88 (0.86–0.89)	0.89	0.86	0.88	0.88 (0.86–0.89)	0.89	0.15
Light GBM	0.66 (0.61–0.71)	0.63	0.61	0.60	0.62 (0.57–0.66)	0.63	0.48
NN	0.68 (0.64–0.73)	0.68	0.60	0.63	0.63 (0.59–0.68)	0.68	0.52

**Abbreviations:** ML, machine learning; CLB, cumulate live birth; AUC, area under the curves; LR, logistic regress; SVM, support vector machine; RF, random forest; XGBoost, extreme gradient boosting; lightGBM, light gradient boosting machine; NN, neural networks.

## Interpretability Analysis and Clinical Translation

The SHAP method was implemented to assess the feature importance of each variable. The comprehensive swarm plots illustrating the variables of the XGBoost model-1 and model-2 were presented in Figure 5A and C. In model-1, SHAP plots demonstrated that the number of 2 PN zygotes, maternal age and the endometrial thickness on trigger day had the broadest range among patients, highlighting their strong predictive power (Figure 5A and B). However, maternal age and the levels of FSH and AMH had the strongest predictive power by SHAP plots in model-2 (Figure 5C and D). Figure 5B and D exhibited the ranking of feature importance of each variable. In model-2, the length of infertility, AFC and treatment of ovarian endometriomas had the similar predictive power, indicating that while endometriomas are a factor, they are less decisive than core markers of ovarian aging in predicting cycle success. These findings suggest that endometriomas may not directly determine ART outcomes, they exert an indirect influence by compromising ovarian reserve. These findings have direct clinical implications. For clinicians using Model-2 prior to an ART cycle, the dominant influence of age and basal FSH/AMH suggests that ovarian reserve assessment is key. A patient with advanced age and diminished reserve (high FSH, low AMH) would receive a low probability score, suggesting the need to discuss: 1) the relative risks and benefits of natural conception versus ART after surgery combining with EFI, 2) whether more individual stimulation protocols might be indicated, and 3) expected oocyte yield and the potential need for multiple cycles.

Model-1 relies primarily on the number of 2PN zygotes and endometrial thickness. A low score driven by poor zygote yield suggests a fertilization or embryo quality issue. This could guide several clinical decisions. If ICSI was not used in the current cycle, it may be worth considering in future attempts. PGT could also help select viable embryos. Alternatively, if endometrial receptivity is a concern, a “freeze-all” strategy might be preferable. Conversely, a low score primarily due to thin



**Figure 5** SHAP interpretation of the XGBoost model. **(A)** The swarm plots of the XGBoost model for live birth prediction based on model-1 with 4 variables. Every features impact on the model's output. **(B)** Feature importance created by XGBoost algorithm in model-1. **(C)** The swarm plots of the XGBoost model for live birth prediction based on model-2 with 9 variables. **(D)** Feature importance created by XGBoost algorithm in model-2. For A and C, every dot in a row symbolizes a patient and its color denotes the feature value – yellow denotes a value that is greater and purple denotes a value that is low. The horizontal axis represented SHAP values, while the vertical axis displayed features sorted by their cumulative SHAP value impact. Each data pointed corresponds to a specific instance, with its position along the x-axis indicating the SHAP value for the particular instance and feature.

**Abbreviations:** PN2, 2-pronuclear; EM, endometrial thickness on trigger day; AMH, anti-Müllerian hormone; FSH, follicle-stimulating hormone; TFC, total antral follicle count; LH, luteinizing hormone; BMI, body mass index; E<sub>2</sub>, estradiol; years, length of infertility.

endometrium shifts focus on optimizing the endometrial environment through adjusted estrogen supplementation or investigating underlying pathologies before proceeding with embryo transfer.

## Discussion

We developed and validated non-linear machine learning models to predict CLB following a complete ART cycle in women with ovarian endometriomas, while also comparing the performance of seven distinct ML algorithms. Employing these algorithms, we trained two predictive models by screening 20 variables including demographics and features associated with endometriomas, ovarian reserve and ART. The model-1 contains 4 features with variables relevant to ART and model-2 contains 9 features without ART variables. Results indicated that the XGBoost algorithm performed robustly in both models, showing strong discrimination and calibration. Both models were further validated using an external cohort, confirming their stability and accuracy. To gain deeper insights into the ML models, the SHAP method was utilized for visualization and revealed that maternal age and characteristics associated with ovarian reserve had the most significant influence on the predictable (model-1: the number of 2 PN zygotes, age and endometrial thickness on trigger day; model-2: age, the levels of FSH and AMH).

Machine learning techniques have been increasingly applied to address complex medical challenges in recent years.<sup>19</sup> While most existing IVF prediction models are built on general infertility cohorts using endpoints such as pregnancy per transfer,<sup>18,24</sup> clinical decision-making for women with ovarian endometriomas requires more nuanced consideration due to the potential impact of cysts on ovarian response and surgical considerations. We therefore developed models specifically for this population, using cumulative live birth rate per ovarian retrieval as the endpoint.

Unlike previous studies that predominantly relied on logistic regression,<sup>18</sup> we employed XGBoost as a non-linear ensemble algorithm that combines multiple weak classifiers and is relatively robust to outliers,<sup>35</sup> enabling it to capture complex interactions among predictors, such as cyst characteristics and prior treatment history, that conventional linear models might miss.<sup>20,24,36,37</sup> The factors affecting ART outcomes are diverse, complex and heterogeneous, such as maternal age, female obesity, ovarian reserve, and the number of oocytes retrieved.<sup>38–42</sup> Given that some of these factors are not linearly related to cumulative live birth, XGBoost, which has demonstrated superior performance among seven algorithms, offers a more accurate predictive approach. Its utility has been increasingly recognized in other complex medical domains such as sepsis, cardiovascular disease, and kidney injury.<sup>43–45</sup>

Most previous studies only compared the IVF/ICSI outcomes among patients with endometriosis but have not developed an eligible predictive model. These studies reported the endometriomas did not adversely affect the likelihood of conception following IVF/ICSI, even though women with endometriomas have lower oocytes.<sup>46–49</sup> Besides, the same conclusions were found between women with surgical treatment of endometriomas and those who refused operation.<sup>48,50,51</sup> For patients, what matters more is when a successful live birth will be achieved, rather than merely the outcomes following a single transplantation. In recent years, cumulative live birth rate has been recommended as a key patient-centered outcome in infertility treatment trials<sup>26</sup> and CLB per retrieval cycle represents a logical extension of this endpoint.<sup>25</sup> Although most endometriomas are unilateral, bilateral cases often correlate with more severe disease and posterior cul-de-sac obliteration.<sup>29</sup> In our study, the endometriomas negatively affected the CLB following ART, especially for patients underwent bilateral cystectomy and ablation, which is consistent with the above researches.<sup>52–54</sup>

This is the first time to develop non-linear ML models to predict the CLB in an ART cycle by compared 7 ML algorithms among patients with endometriomas. The model-1 and 2 was used to predict the likelihood of live birth before and after ART separately. The model-2 is mainly used to enhance reproductive strategies after surgical removing of lesions and model-1 is expected to assisted embryo transfer strategies after retrieving oocytes.

The recommended management of endometriosis-associated infertility includes ART, surgery, and expectant management. The decision to perform surgery should consider pain symptoms, patient age and preferences, prior surgical history, other infertility factors, ovarian reserve, and the Endometriosis Fertility Index (EFI).<sup>55</sup> It is difficult to formulate a recommendation for patients with endometriomas. Our ML Model-2, particularly when used in combination with the EFI, may help identify patients who would benefit most from ART.

Our models may serve as a supportive tool to inform, but not replace, clinical decision-making at two stages. Model-2, based on pre-ART variables, can inform the initial consultation: a high predicted probability may favor proceeding directly to

ART over expectant management or repeat surgery, particularly when considered alongside tools like the EFI. In contrast, model-1, which incorporates post-retrieval parameters, helps support embryo transfer decisions. A low probability might suggest discussing options such as a “freeze-all” cycle, personalized luteal-phase support,<sup>56</sup> or realistic expectations for the current cycle.

SHAP improves the interpretability of ML predictions by illustrating the contribution of each feature. Consistent to the previous researches,<sup>25,57,58</sup> the women age and variables related to ovarian reserve (AMH/FSH/AFC) had the strongest power to predict CLB following a single ART cycle. The ovarian endometriomas demonstrated a certain predictive capability for CLB in both model-1 and model-2. Prior studies have shown that women with endometriomas have less healthy ovarian tissue than those with other benign cysts.<sup>59</sup> Chronic inflammation associated with endometriomas may further impair ovarian, tubal, or endometrial function, leading to abnormalities in folliculogenesis, fertilization, or implantation.<sup>60</sup> Additionally, early follicles in ovaries with endometriomas are more prone to atresia, thereby diminishing ovarian reserve.<sup>61</sup> Among patients with unilateral endometriomas, the affected ovary often shows reduced response to ovarian stimulation compared to the contralateral ovary.<sup>62</sup> These studies verified that the ovarian endometriomas could reduce the probability of obtaining live births by impairing ovarian reserve.

This is the first study to develop and externally validate non-linear machine learning models specifically to predict cumulative live birth following ART in women with ovarian endometriomas. After comparing seven algorithms, we created two clinically tailored models—one for pre-ART counseling and one for post-retrieval decisions—to support different stages of the treatment pathway. The primary strength of our predictive models is first used XGBoost algorithms among patients with endometriomas, which can effectively capture non-linear relationships and have demonstrated considerable predictive performance in predicting CLB. Furthermore, the variables included in Model-2 are routinely available in clinical practice, facilitating potential widespread adoption.

This study has several limitations. The retrospective design introduces potential selection bias, information bias, and unmeasured confounders. We used pairwise deletion for missing data (which were below 1%) and while our diagnostic checks supported the missing completely at random (MCAR) assumption, we cannot definitively rule out non-random missingness. Similarly, although we applied SMOTE to effectively address class imbalance and improved model sensitivity, it may affect probability calibration. Although the calibration assessment suggested acceptable performance, clinicians should be aware that predicted probabilities may not perfectly reflect absolute risk. Our analysis was also constrained by unavailable data on certain ovarian reserve markers and endometrioma characteristics prior to surgery. Access to these variables could improve model performance and help disentangle the effects of the cysts themselves from those of prior surgical treatment. Additionally, because our cohort was drawn from two tertiary centers within the same university system, the findings may not generalize to other settings. While we performed external validation, both centers serve similar populations, limiting geographic diversity and practice patterns. Prospective, multi-center studies are needed to validate their performance and clinical impact. Even so, our findings suggest that ML models can help personalize reproductive strategy selection for women with endometriomas.

## Conclusion

We developed two XGBoost-based models to predict cumulative live birth following ART in women with ovarian endometriomas. Both performed reasonably well. Model-1, which incorporated post-retrieval variables, was more accurate than Model-2 (which relied solely on pre-ART data). SHAP analysis identified maternal age and ovarian reserve markers as the dominant predictors; prior intervention for endometriomas had a relatively modest influence. Clinically, these tools could support counseling at two key decision points—initial treatment strategy and embryo transfer management.

However, the models were built on retrospective data from a single health system, so generalizability remains uncertain. Prospective validation in multi-center settings will be critical before broader adoption. We are therefore planning to host the models on a web platform to facilitate external testing and real-world evaluation.

## Ethics Approval Statement

The study was approved by the Reproductive Medicine Ethics Committee, Hospital for Reproductive Medicine Affiliated to Shandong University. All parents consented to the gathering of their data.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; they gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This study was partly funded by the National Key Technology Research and Developmental Program of China (2024YFC2706902), the National Key R&D Program of China (2022YFC2704400, 2024YFC2706700), the Key R&D Program of Shandong Province, China (2025CXGC020308), the National Natural Science Foundation of China (NSFC) Regional Innovation and Development Joint Fund (U24A20664), CAMS Innovation Fund for Medical Sciences (2021-I2M-5-001), the Excellence Research Group Program of NSFC (32588201), and National Special Support Program for High-level Talents.

## Disclosure

All authors have declared no conflicts of interest in this work.

## References

- Zondervan KT, Becker CM, Koga K, et al. Endometriosis. *Nat Rev Dis Primers*. 2018;4(1):9. doi:10.1038/s41572-018-0008-5
- Taylor HS, Kotlyar AM, Flores VA. Endometriosis is a chronic systemic disease: clinical challenges and novel innovations. *Lancet*. 2021;397(10276):839–852. doi:10.1016/s0140-6736(21)00389-5
- Muzii L, Di Tucci C, Di Feliciano M, et al. Management of endometriomas. *Semin Reprod Med*. 2017;35(1):25–30. doi:10.1055/s-0036-1597126
- Matsuzaki S, Schubert B. Oxidative stress status in normal ovarian cortex surrounding ovarian endometriosis. *Fertil Steril*. 2010;93(7):2431–2432. doi:10.1016/j.fertnstert.2009.08.068
- Agarwal A, Aponte-Mellado A, Premkumar BJ, et al. The effects of oxidative stress on female reproduction: a review. *Reprod Biol Endocrinol*. 2012;10(1):49. doi:10.1186/1477-7827-10-49
- Medicine PCotASIR. Endometriosis and infertility: a committee opinion. *Fertil Steril*. 2012;98(3):591–598. doi:10.1016/j.fertnstert.2012.05.031
- Chapron C, Marcellin L, Borghese B, Santulli P. Rethinking mechanisms, diagnosis and management of endometriosis. *Nat Rev Endocrinol*. 2019;15(11):666–682. doi:10.1038/s41574-019-0245-z
- Khine YM, Taniguchi F, Harada T. Clinical management of endometriosis-associated infertility. *Reprod Med Biol*. 2016;15(4):217–225. doi:10.1007/s12522-016-0237-9
- Vercellini P, Somigliana E, Viganò P, et al. Surgery for endometriosis-associated infertility: a pragmatic approach. *Hum Reprod*. 2009;24(2):254–269. doi:10.1093/humrep/den379
- Goodman LR, Goldberg JM, Flyckt RL, et al. Effect of surgery on ovarian reserve in women with endometriomas, endometriosis and controls. *Am J Obstet Gynecol*. 2016;215(5):589.e581–589.e586. doi:10.1016/j.ajog.2016.05.029
- Raffi F, Metwally M, Amer S. The impact of excision of ovarian endometrioma on ovarian reserve: a systematic review and meta-analysis. *J Clin Endocrinol Metab*. 2012;97(9):3146–3154. doi:10.1210/jc.2012-1558
- Hirokawa W, Iwase A, Goto M, et al. The post-operative decline in serum anti-Müllerian hormone correlates with the bilaterality and severity of endometriosis. *Hum Reprod*. 2011;26(4):904–910. doi:10.1093/humrep/der006
- Mehdizadeh Kashi A, Chaichian S, Ariana S, et al. The impact of laparoscopic cystectomy on ovarian reserve in patients with unilateral and bilateral endometrioma. *Int J Gynaecol Obstet*. 2017;136(2):200–204. doi:10.1002/ijgo.12046
- Celik HG, Dogan E, Okay E, et al. Effect of laparoscopic excision of endometriomas on ovarian reserve: serial changes in the serum antimüllerian hormone levels. *Fertil Steril*. 2012;97(6):1472–1478. doi:10.1016/j.fertnstert.2012.03.027
- Adamson GD, Pasta DJ. Endometriosis fertility index: the new, validated endometriosis staging system. *Fertil Steril*. 2010;94(5):1609–1615. doi:10.1016/j.fertnstert.2009.09.035
- Tomassetti C, Bafort C, Meuleman C, et al. Reproducibility of the endometriosis fertility index: a prospective inter-/intra-rater agreement study. *Bjog*. 2020;127(1):107–114. doi:10.1111/1471-0528.15880
- Vesali S, Razavi M, Rezaeinejad M, et al. Endometriosis fertility index for predicting non-assisted reproductive technology pregnancy after endometriosis surgery: a systematic review and meta-analysis. *Bjog*. 2020;127(7):800–809. doi:10.1111/1471-0528.16107
- Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Hum Reprod*. 2020;35(1):100–116. doi:10.1093/humrep/dez258
- Shehab M, Abualigah L, Shambour Q, et al. Machine learning in medical applications: a review of state-of-the-art methods. *Comput Biol Med*. 2022;145:105458. doi:10.1016/j.combiomed.2022.105458
- Barnett-Itzhaki Z, Elbaz M, Buttermann R, et al. Machine learning vs. classic statistics for the prediction of IVF outcomes. *J Assist Reprod Genet*. 2020;37(10):2405–2412. doi:10.1007/s10815-020-01908-1
- Sadegh-Zadeh SA, Khanjani S, Javanmardi S, et al. Catalyzing IVF outcome prediction: exploring advanced machine learning paradigms for enhanced success rate prognostication. *Front Artif Intell*. 2024;7:1392611. doi:10.3389/frai.2024.1392611

22. Illingworth PJ, Venetis C, Gardner DK, et al. Deep learning versus manual morphology-based embryo selection in IVF: a randomized, double-blind noninferiority trial. *Nat Med.* 2024;30(11):3114–3120. doi:10.1038/s41591-024-03166-5
23. Peng J, Geng X, Zhao Y, et al. Machine learning algorithms in constructing prediction models for assisted reproductive technology (ART) related live birth outcomes. *Sci Rep.* 2024;14(1):32083. doi:10.1038/s41598-024-83781-x
24. Dhillon RK, McLernon DJ, Smith PP, et al. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Hum Reprod.* 2016;31(1):84–92. doi:10.1093/humrep/dev268
25. Yan J, Qin Y, Zhao H, et al. Live birth with or without preimplantation genetic testing for aneuploidy. *N Engl J Med.* 2021;385(22):2047–2058. doi:10.1056/NEJMoa2103613
26. Duffy JMN, Bhattacharya S, Bhattacharya S, et al. Standardizing definitions and reporting guidelines for the infertility core outcome set: an international consensus development study. *Fertil Steril.* 2021;115(1):201–212. doi:10.1016/j.fertnstert.2020.11.013
27. Singh SS, Suen MW. Surgery for endometriosis: beyond medical therapies. *Fertil Steril.* 2017;107(3):549–554. doi:10.1016/j.fertnstert.2017.01.001
28. Mansouri G, Safinataj M, Shahesmaeili A, et al. Effect of laparoscopic cystectomy on ovarian reserve in patients with ovarian cyst. *Front Endocrinol.* 2022;13:964229. doi:10.3389/fendo.2022.964229
29. Baraki D, Richards EG, Falcone T. Treatment of endometriomas: surgical approaches and the impact on ovarian reserve, recurrence, and spontaneous pregnancy. *Best Pract Res Clin Obstet Gynaecol.* 2024;92:102449. doi:10.1016/j.bpobgyn.2023.102449
30. Liu XZ, Duan M, Huang HD, et al. Predicting diabetic kidney disease for type 2 diabetes mellitus by machine learning in the real world: a multicenter retrospective study. *Front Endocrinol.* 2023;14:1184190. doi:10.3389/fendo.2023.1184190
31. Liu X, Xie Z, Zhang Y, et al. Machine learning for predicting in-hospital mortality in elderly patients with heart failure combined with hypertension: a multicenter retrospective study. *Cardiovasc Diabetol.* 2024;23(1):407. doi:10.1186/s12933-024-02503-9
32. Wang K, Tian J, Zheng C, et al. Improving risk identification of adverse outcomes in chronic heart failure using SMOTE+ENN and machine learning. *Risk Manag Healthc Policy.* 2021;14:2453–2463. doi:10.2147/rmhp.S310295
33. Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning.* 2024;113(7):4903–4923. doi:10.1007/s10994-022-06296-4
34. Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Fail.* 2020;8(1):12–21. doi:10.1016/j.jchf.2019.06.013
35. Guan C, Gong A, Zhao Y, et al. Interpretable machine learning model for new-onset atrial fibrillation prediction in critically ill patients: a multi-center study. *Crit Care.* 2024;28(1):349. doi:10.1186/s13054-024-05138-0
36. Salih M, Austin C, Warty RR, et al. Embryo selection through artificial intelligence versus embryologists: a systematic review. *Hum Reprod Open.* 2023;2023(3):hoad031. doi:10.1093/hropen/hoad031
37. Jiang VS, Bormann CL. Artificial intelligence in the in vitro fertilization laboratory: a review of advancements over the last decade. *Fertil Steril.* 2023;120(1):17–23. doi:10.1016/j.fertnstert.2023.05.149
38. Zhou QW, Jing S, Xu L, et al. Clinical and neonatal outcomes of patients of different ages following transfer of thawed cleavage embryos and blastocysts cultured from thawed cleavage-stage embryos. *PLoS One.* 2018;13(11):e0207340. doi:10.1371/journal.pone.0207340
39. Ding W, Zhang FL, Liu XC, et al. Impact of female obesity on cumulative live birth rates in the first complete ovarian stimulation cycle. *Front Endocrinol.* 2019;10:516. doi:10.3389/fendo.2019.00516
40. Lew R. Natural history of ovarian function including assessment of ovarian reserve and premature ovarian failure. *Best Pract Res Clin Obstet Gynaecol.* 2019;55:2–13. doi:10.1016/j.bpobgyn.2018.05.005
41. Tal R, Seifer DB, Tal R, et al. AMH highly correlates with cumulative live birth rate in women with diminished ovarian reserve independent of age. *J Clin Endocrinol Metab.* 2021;106(9):2754–2766. doi:10.1210/clinem/dgab168
42. McLernon DJ, Raja EA, Toner JP, et al. Predicting personalized cumulative live birth following in vitro fertilization. *Fertil Steril.* 2022;117(2):326–338. doi:10.1016/j.fertnstert.2021.09.015
43. Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGBoost. *J Transl Med.* 2020;18(1):462. doi:10.1186/s12967-020-02620-5
44. Li J, Liu S, Hu Y, et al. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. *J Med Internet Res.* 2022;24(8):e38082. doi:10.2196/38082
45. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care.* 2019;23(1):112. doi:10.1186/s13054-019-2411-z
46. Hamdan M, Omar SZ, Dunselman G, Cheong Y. Influence of endometriosis on assisted reproductive technology outcomes: a systematic review and meta-analysis. *Obstet Gynecol.* 2015;125(1):79–88. doi:10.1097/aog.0000000000000592
47. Harb HM, Gallos ID, Chu J, et al. The effect of endometriosis on in vitro fertilisation outcome: a systematic review and meta-analysis. *Bjog.* 2013;120(11):1308–1320. doi:10.1111/1471-0528.12366
48. Hamdan M, Dunselman G, Li TC, Cheong Y. The impact of endometrioma on IVF/ICSI outcomes: a systematic review and meta-analysis. *Hum Reprod Update.* 2015;21(6):809–825. doi:10.1093/humupd/dmv035
49. Wu Y, Yang R, Lan J, et al. Ovarian endometrioma negatively impacts oocyte quality and quantity but not pregnancy outcomes in women undergoing IVF/ICSI treatment: a retrospective cohort study. *Front Endocrinol.* 2021;12:739228. doi:10.3389/fendo.2021.739228
50. Collinet P, Fritel X, Revel-Delhom C, et al. Management of endometriosis: CNGOF/HAS clinical practice guidelines - Short version. *J Gynecol Obstet Hum Reprod.* 2018;47(7):265–274. doi:10.1016/j.jogoh.2018.06.003
51. Paik H, Jee BC. Impact of ablation versus cystectomy for endometrioma on ovarian reserve, recurrence, and pregnancy: an updated meta-analysis. *Reprod Sci.* 2024;31(7):1924–1935. doi:10.1007/s43032-024-01512-z
52. Horton J, Sterrenburg M, Lane S, et al. Reproductive, obstetric, and perinatal outcomes of women with adenomyosis and endometriosis: a systematic review and meta-analysis. *Hum Reprod Update.* 2019;25(5):592–632. doi:10.1093/humupd/dmz012
53. Bourdon M, Mimouni A, Maignien C, et al. Reduced live birth rates following ART in adenomyosis patients: a matched control study. *Hum Reprod.* 2025;40(5):855–864. doi:10.1093/humrep/deaf052
54. Alson S, Stenqvist A, Sladkevicius P. Cumulative live birth rates under three consecutive IVF/ICSI treatment cycles are reduced in women with endometriosis and/or adenomyosis diagnosed by ultrasonography. *Hum Reprod.* 2025;40(12):2332–2341. doi:10.1093/humrep/deaf184
55. Becker CM, Bokor A, Heikinheimo O, et al. ESHRE guideline: endometriosis. *Hum Reprod Open.* 2022;2022(2):hoac009. doi:10.1093/hropen/hoac009

56. Cimadomo D, de Los Santos MJ, Griesinger G, et al. ESHRE good practice recommendations on recurrent implantation failure. *Hum Reprod Open.* 2023;2023(3):hoad023. doi:10.1093/hropen/hoad023
57. Peigné M, Bernard V, Dijols L, et al. Using serum anti-Müllerian hormone levels to predict the chance of live birth after spontaneous or assisted conception: a systematic review and meta-analysis. *Hum Reprod.* 2023;38(9):1789–1806. doi:10.1093/humrep/dead147
58. Xia L, Zhou X, Wang X, et al. The role of age and AMH on cumulative live birth rates over multiple frozen-thawed embryo transfer cycles: a study based on low prognosis patients of POSEIDON 3 and 4 groups. *Reprod Biol Endocrinol.* 2024;22(1):69. doi:10.1186/s12958-024-01243-5
59. Maneschi F, Marasá L, Incandela S, et al. Ovarian cortex surrounding benign neoplasms: a histologic study. *Am J Obstet Gynecol.* 1993;169(2):388–393. doi:10.1016/0002-9378(93)90093-x
60. Macer ML, Taylor HS. Endometriosis and infertility: a review of the pathogenesis and treatment of endometriosis-associated infertility. *Obstet Gynecol Clin North Am.* 2012;39(4):535–549. doi:10.1016/j.ogc.2012.10.002
61. Kitajima M, Dolmans MM, Donnez O, et al. Enhanced follicular recruitment and atresia in cortex derived from ovaries with endometriomas. *Fertil Steril.* 2014;101(4):1031–1037. doi:10.1016/j.fertnstert.2013.12.049
62. Coccia ME, Rizzello F, Barone S, et al. Is there a critical endometrioma size associated with reduced ovarian responsiveness in assisted reproduction techniques? *Reprod Biomed Online.* 2014;29(2):259–266. doi:10.1016/j.rbmo.2014.04.019

## Clinical Epidemiology

### Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

**Dovepress**

Taylor & Francis Group