

RESEARCH

Open Access



Construction and evaluation of machine learning-based prediction model for live birth following fresh embryo transfer in IVF/ICSI patients with polycystic ovary syndrome

Suqin Zhu^{1,2†}, Zhiqing Huang^{1,3†}, Xiaojing Chen¹, Wenwen Jiang¹, Yuan Zhou¹, Beihong Zheng^{1*} and Yan Sun^{1,3*}

Abstract

Objective To investigate the determinants affecting live birth outcomes in fresh embryo transfer among polycystic ovary syndrome (PCOS) patients using various machine learning (ML) algorithms and to construct predictive models, offering novel insights for enhancing live birth rates in this specific group.

Methods A sum of 1,062 fresh embryo transfer cycles involving PCOS patients were analyzed, with 466 resulting in live births. The dataset was split randomly into training and testing subsets at a 7:3 ratio. Least absolute shrinkage and selection operator and recursive feature elimination methods were utilized for feature selection within the training data. A grid search strategy identified the optimal parameters for seven ML models: decision tree (DT), K-nearest neighbors (KNN), light gradient boosting machine (LightGBM), naive Bayes model (NBM), random forest (RF), support vector machine (SVM) and extreme gradient boosting (XGBoost). The evaluation of model effectiveness incorporated diverse metrics, encompassing area under the curve (AUC), accuracy, positive predictive value, negative predictive value, F1 score, and Brier score. Calibration curves and decision curve analysis were employed to ascertain the optimal model. Furthermore, Shapley additive explanations were applied to elucidate the importance of predictor variables in the top-performing model.

Results The AUC values of DT, KNN, LightGBM, NBM, RF, SVM and XGBoost models in the training set were 0.813, 1.000, 0.724, 0.791, 1.000, 0.819 and 0.853, respectively. Corresponding values in the testing set were 0.773, 0.719, 0.705, 0.764, 0.794, 0.806 and 0.822. XGBoost emerged as the most effective ML model. SHAP analysis revealed that variables encompassing embryo transfer count, embryo type, maternal age, infertility duration, body mass index, serum testosterone (T) levels, and progesterone (P) levels on the day of human chorionic gonadotropin administration were pivotal predictors of live birth outcomes in individuals with PCOS receiving fresh embryo transfer.

[†]Suqin Zhu and Zhiqing Huang contributed equally to this work.

*Correspondence:

Beihong Zheng
zhengbeihong2010@163.com

Yan Sun
sunyanteam@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusion This study developed a live birth prediction model tailored for PCOS fresh embryo transfer cycles, leveraging ML algorithms to compare the efficacy of multiple models. The XGBoost model demonstrated superior predictive capacity, enabling prompt and precise identification of critical risk factors influencing live birth outcomes in PCOS patients. These findings offer actionable insights for clinical intervention, guiding strategies to improve pregnancy outcomes in this population.

Clinical trial number Not applicable.

Keywords Fresh embryo transfer, Live birth, Machine learning, Polycystic ovary syndrome

Polycystic ovary syndrome (PCOS) is widely recognized as a prevalent reproductive, endocrine, and metabolic disorder impacting women of childbearing age. Beyond pathophysiological irregularities—such as ovulatory dysfunction, infertility, hyperandrogenism (HA), excessive body weight, hyperinsulinemia, and metabolic syndrome—psychological challenges, including anxiety and depression, also influence reproductive well-being. These individuals not only encounter difficulties in achieving conception but are also exposed to heightened risks of miscarriage and preterm delivery during pregnancy. Additionally, pregnancy-related complications and comorbidities, such as gestational diabetes mellitus (GDM) and hypertensive disorders of pregnancy, exhibit a marked increase [1, 2], with potential long-term implications for the metabolic health and development of offspring [1, 3].

In vitro fertilization-embryo transfer (IVF-ET) technology, along with its derivative method, intracytoplasmic sperm injection (ICSI), has been extensively applied in the treatment of infertility, with the proportion of PCOS patients steadily rising over the years [4]. The advancement of machine learning (ML) algorithms has facilitated the development of numerous ML-based models within the domain of assisted reproduction [5]. These algorithms offer remarkable adaptability in handling intricate nonlinear relationships among variables and identifying complex data patterns and interactions, which allows for more precise predictions and enhanced model performance [6]. Moreover, in comparison with logistic regression, ML approaches exhibit superior robustness against outliers, reduced sensitivity to extreme values, and more effective data-processing capabilities, whereas the susceptibility of logistic regression to outliers may result in skewed outcomes [7].

With the advancement of ML applications in the medical field, considerable strides have been achieved in the early diagnosis and management of PCOS [8, 9]. Artificial intelligence-driven techniques are capable of recognizing patterns within medical data, such as hormonal profiles, facilitating the differentiation between PCOS and non-PCOS individuals. This enhanced precision allows for earlier and more accurate diagnoses, ultimately contributing to improved pregnancy outcomes for patients with PCOS [10]. Although previous studies

have investigated factors influencing live births following IVF treatment among PCOS patients [11–12], these studies are based on the frozen-thawed embryo transfer cycles, with a deficiency in the analysis of the influencing factors of the outcomes of fresh embryo transfer cycles in PCOS. In recent years, through effective improve the body metabolism and endocrine abnormalities, and accept low dose gonadotropin (GN) stimulate ovulation treatment, more and more patients with PCOS were offered the opportunity to fresh embryo transfer. For this group of people, our study is useful for the targeted interventions and the development of personalized treatment. Moreover, research focusing on the use of ML for predicting live births within this context remains scarce. The purpose of this investigation is to explore influential factors contributing to live births after fresh embryo transfer in individuals with PCOS, determine reliable predictive indicators, and construct prediction models for live birth outcomes after fresh embryo transfer. Seven ML models were developed and assessed using demographic, biochemical, and embryological data, and their predictive performance characteristics were compared. Furthermore, Shapley additive explanations (SHAP) were utilized to analyze and interpret each predictor's contribution, addressing the challenges associated with the interpretability of ML models and providing clinical guidance for early intervention in this population.

Methodologies and materials

General information

This retrospective cohort study investigated assisted reproductive populations, focusing specifically on female PCOS patients who underwent the antagonist protocol followed by fresh embryo transfer at the Fujian Provincial Maternal and Child Health Hospital between January 2019 and December 2023.

Inclusion criteria: ① PCOS patients meeting the Rotterdam diagnostic criteria [13] or Chinese guidelines for PCOS diagnosis and treatment [14]; ② ovarian stimulation with antagonist protocol; ③ fresh embryo transfer cycles.

Exclusion criteria: ① uterine abnormalities including uterine malformation, adenomyosis, and submucosal fibroids; ② endometriosis; ③ hydrosalpinx; ④

chromosomal abnormalities in either partner; ⑤ severe oligoasthenozoospermia in male partners; ⑥ loss to follow-up or missing outcome data. This investigation per the principles of the Declaration of Helsinki.

Ovarian stimulation and *in vitro* fertilization (IVF)

All patients underwent ultrasound assessments and hormone assays on the second or third day of their menstrual cycle. Ovarian stimulation was initiated with Gn [Gonal-F, recombinant FSH (rFSH), Merck Serono; recombinant follitropin beta injection (rFSH), MSD (China); or urofollitropin for injection, Livzon Pharmaceutical Group] at an initial dosage ranging from 112.5 to 225 U. GnRH antagonists (Ganirelix, 0.25 mg, Organon; or Cetrotide, 0.25 mg, Merck Serono) were administered once follicles exceeded 12 mm in diameter and/or serum estradiol concentrations reached at least 1,000 pmol/L. Dosage adjustments were made based on each patient's response. When two follicles measured 18 mm or more, or three follicles reached at least 17 mm in diameter, final oocyte maturation was triggered with either 250 µg recombinant human chorionic gonadotropin (Ovidrel, Merck Serono, Switzerland) or a dual trigger involving 0.2 mg triptorelin acetate (Decapeptyl, Ferring, Germany) administered subcutaneously, combined with 2,000 IU human chorionic gonadotropin (HCG, Livzon, China) administered intramuscularly.

Oocyte retrieval was conducted under the guidance of transvaginal ultrasound 36–38 h following the trigger. Transfer of either a fresh cleavage-stage embryo or a blastocyst was performed on the third or fifth-day post-retrieval. Standard luteal phase support was initiated after oocyte retrieval, comprising oral dydrogesterone tablets (Duphaston, Abbott, Netherlands), administered every eight hours, along with an progesterone vaginal sustained-release gel (Crinone, Merck Serono, Switzerland) at a dosage of 90 mg per day via vaginal administration. A positive hCG result was defined by serum hCG levels exceeding 5 U/L on the 14th day after cleavage-stage embryo transfer or the 12th day following blastocyst transfer. Clinical pregnancy was confirmed by the presence of at least one gestational sac, with or without a fetal pole, located within or outside the uterus, as verified through transvaginal ultrasound.

Data collection and preprocessing

For eligible patients, basic clinical data were collected, including: ① demographic information (such as age, body mass index (BMI), duration of infertility, and number of treatment cycles); ② laboratory test results (e.g., basal follicle-stimulating hormone (FSH), luteinizing hormone, and estradiol (E2) levels); ③ treatment procedures (e.g., GN dosage, number of transferred embryos, and types of transferred embryos). Data cleaning was conducted in

accordance with patterns of missing values. Initially, rows containing more than 20% missing data across all samples were excluded. Subsequently, columns with over 20% missing values across all rows were also eliminated. Any remaining missing values were imputed using the `missForest` function in R.

Data feature definition

A live birth was characterized as a pregnancy reaching 28 weeks or more, resulting in the presence of at least one of the four vital signs following delivery: heartbeat, respiration, umbilical cord pulsation, or muscle tone.

Statistical analysis

Python 3.9, R 4.3.2, and SPSS 26.0 were utilized to perform data analysis. Continuous variables were evaluated utilizing either the t-test or the Mann-Whitney U test, while categorical variables were analyzed with the chi-square test. Continuous data were expressed as median values, whereas categorical data were described as frequencies or percentages. The significance level (α) was established at 0.05. The dataset was split arbitrarily into training and testing subsets at a ratio of 7:3. Least absolute shrinkage and selection operator (LASSO) regression, in combination with recursive feature elimination (RFE), was utilized to identify predictive factors, thereby enhancing both predictive performance and model interpretability. Optimal parameters for multiple ML models were ascertained utilizing five-fold cross-validation and grid search. Decision tree, K-nearest neighbors, light gradient boosting machine, naive Bayes model, random forest, support vector machine, and extreme gradient boosting (XGBoost) models were developed using five-fold cross-validation. The discriminative ability of these models in predicting live birth events was evaluated through metrics encompassing area under the curve (AUC), accuracy, precision, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Brier score. Calibration curves were employed to assess the alignment between forecasted probabilities and actual outcomes in the optimal model. SHAP was applied to explain the optimal model. The statistical analysis and model development workflow are illustrated in Fig. 1.

Results

Patient characteristics

A sum of 1,062 cycles was analyzed in this investigation, with 466 resulting in live births. The modeling group comprised 743 cycles, while the validation group included 319 cycles. Furthermore, no statistically significant differences were detected in the baseline data and clinical characteristics between these two groups (all $P > 0.05$), affirming their comparability (Table 1). Table 2 depicts a comparison of patient characteristics between

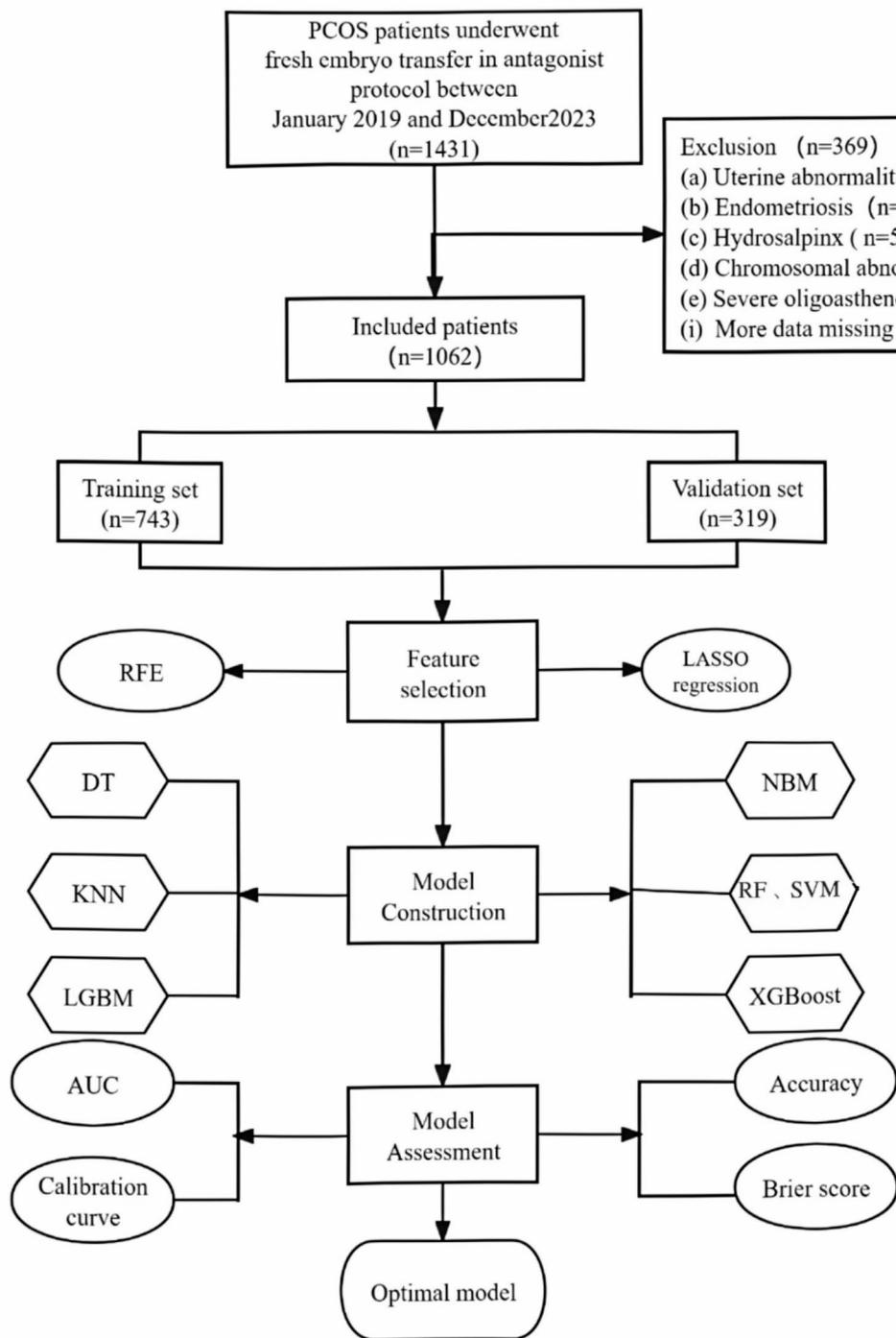


Fig. 1 Flow chart of the study

the live birth group and the control group. Significant differences ($P < 0.05$) were identified for maternal age, BMI, infertility duration, treatment frequency, serum testosterone (T) levels, initial GN dosage, FSH levels on the day of HCG administration, progesterone (P) levels on the HCG day, the number of high-quality cleavage-stage

embryos, the type of transferred embryos, and the number of transferred embryos between the groups.

Selection of predictive factors

LASSO regression, in combination with RFE, was employed to identify predictive factors. Figure 2A and

Table 1 Basis characteristics of training group and validation group

Characteristics	ALL(n= 1062)	Training group (n=743)	Validation group (n= 319)	P
Female age (year)	31 (28, 33)	31 (28, 33)	31 (28, 33)	0.828
Male age (year)	32 (30, 35)	32 (30, 35)	32 (30, 35)	0.889
BMI(kg/m ²), n(%)				0.316
<28	889 (83.71)	628 (84.52)	261(81.82)	
≥ 28	173 (16.29)	115 (15.48)	58(18.18)	
Infertility duration (year), n(%)				0.295
<3	477 (44.92)	342 (46.03)	135 (42.32)	
≥ 3	585 (55.08)	401 (53.97)	184 (57.68)	
Number of treatments	1 (1, 3)	1 (1, 3)	1 (1, 3)	0.195
Basal FSH (IU/L)	5.72 (4.88, 6.62)	5.74 (4.90, 6.62)	5.69 (4.88, 6.65)	0.529
Basal LH (IU/L)	6.55 (5.55, 7.95)	6.45 (5.55, 7.85)	6.65 (5.61, 8.25)	0.149
Basal PRL (ng/mL)	14.90 (10.70, 21.10)	14.50 (10.60, 20.80)	15.50 (11.30, 21.95)	0.082
Basal E ₂ (pg/mL)	29.00 (23.00, 39.00)	29.00 (23.00, 39.00)	29.00 (22.00, 39.00)	0.587
T (ng/mL)	0.40 (0.34, 0.49)	0.41 (0.34, 0.50)	0.40 (0.34, 0.49)	0.346
AFC	18 (15, 21)	18 (15, 21)	18 (15, 21)	0.432
AMH (μg/L)	6.74 (5.64, 8.46)	6.71 (5.61, 8.42)	6.82 (5.69, 8.68)	0.359
Fertilization, n(%)				0.737
IVF	788 (74.20)	554 (74.56)	234 (73.35)	
ICSI	274 (25.80)	189 (25.44)	85 (26.65)	
Initial dose of Gn	125 (112.50, 150)	125 (112.50, 150)	125 (112.50, 150)	0.680
Total dose of Gn	2250 (1800.00, 2787.50)	2237.50 (1800.00, 2762.50)	2250.00 (1775.00, 2906.25)	0.577
Dosing days of Gn	12 (11, 14)	12 (11, 14)	12 (11, 14)	0.819
FSH on HCG day (IU/L)	10.56 (8.57, 13.03)	10.62 (8.64, 13.06)	10.46 (8.51, 12.95)	0.946
LH on HCG day (IU/L)	1.00 (0.70, 1.50)	1.00 (0.80, 1.50)	1.04 (0.70, 1.50)	0.551
E2 on HCG day (pg/mL)	3223.00(2263.25, 4259.00)	3267.00 (2314.50, 4326.00)	3143.00 (2200.00, 4203.50)	0.392
P on HCG day (ng/mL)	0.66 (0.46, 0.98)	0.65 (0.47, 0.91)	0.67 (0.44, 1.02)	0.881
Endometrial thickness on HCG day(mm)	11.85 (10.50, 13.00)	11.80 (10.50, 13.00)	12.00 (10.45, 13.00)	0.820
Trigger strategy, n(%)				0.353
rHCG	409 (38.51)	282 (37.95)	127 (39.81)	
GnRH-a combined with HCG	653(61.49)	461 (62.05)	192 (60.19)	
Number of oocytes retrieved	13.00 (10.00, 17.00)	13.00 (10.00, 17.00)	14.00 (11.00, 18.00)	0.153
Number of mature oocytes	8.00 (5.00, 11.00)	8.00 (5.00, 11.00)	7.00 (5.00, 12.00)	0.787
Number of fertilized eggs	10.00 (7.00, 13.00)	10.00 (7.00, 13.00)	10.00 (8.00, 14.00)	0.077
Number of high-quality Cleavage embryos	5.00 (3.00, 8.00)	5.00 (3.00, 8.00)	6.00 (3.00, 9.00)	0.104
Stage of embryos transferred, n(%)				0.540
Cleavage	927 (87.29)	645 (86.81)	282 (88.40)	
Blastocyst	135 (12.71)	98 (13.19)	37 (11.60)	
Number of embryos transferred, n(%)				0.613
1	356 (33.52)	245 (32.97)	111 (34.80)	
2	706 (66.48)	498 (67.03)	208 (65.20)	
Live Birth, n(%)				0.639
yes	466 (43.88)	330 (44.41)	136 (42.63)	
no	596 (56.12)	413 (55.59)	183 (57.37)	

Note: Continuous variables are presented as medians (P25, P75), and categorical data were reported as numbers (%). * $P < 0.05$ was considered statistically significant. BMI: body mass index, FSH: follicle stimulating hormone, LH: luteinizing hormone, E2: Estradiol, T: testosterone, AMH: anti Müllerian hormone, AFC: antral follicle count, Gn: gonadotropin, hCG: human chorionic gonadotropin, P: progesterone, GnRH-a: gonadotropin-releasing hormone agonist

2 C illustrate the variable selection process based on LASSO regression, yielding a feature set of 9 through ten-fold cross-validation, with Lambda 1se serving as the selection criterion. The impact of increasing feature numbers on model accuracy, determined using the RFE approach, is depicted in Fig. 2D and E. The model

attained its highest accuracy when the feature count reached 10. The final feature set is presented in Fig. 2E, where the left circle displays features identified by LASSO regression and the right circle highlights features selected through RFE. The intersection of these two sets

Table 2 Basis characteristics of live birth group and control group

Characteristics	ALL (n = 743)	Live birth group (n = 330)	Control group (n = 413)	P
Female age (year)	31 (28, 33)	30 (28, 32)	32 (29, 34)	< 0.001
Male age (year)	32 (30, 35)	32 (30, 34)	33 (30, 36)	< 0.001
BMI(kg/m ²), n(%)				< 0.001
< 28	628 (84.52)	300 (90.9)	328 (79.42)	
≥ 28	115 (15.48)	30 (9.1)	85 (20.58)	
Infertility duration (year), n(%)				< 0.001
< 3	342 (46.03)	180 (54.55)	162 (39.23)	
≥ 3	401 (53.97)	150 (45.45)	251 (60.77)	
Number of treatments	1 (1, 3)	1 (1, 3)	1 (1, 3)	0.08
Basal FSH (IU/L)	5.74 (4.9, 6.62)	5.68 (4.85, 6.5)	5.84 (4.94, 6.65)	0.155
Basal LH (IU/L)	6.45 (5.55, 7.85)	6.55 (5.65, 7.75)	6.45 (5.55, 7.95)	0.825
Basal PRL (ng/mL)	14.5 (10.6, 20.8)	15.65 (10.9, 21)	13.6 (10.4, 20.4)	0.074
Basal E ₂ (pg/mL)	29 (23, 39)	29 (22.25, 38)	29 (23, 40)	0.506
T (ng/mL)	0.40 (0.34, 0.49)	0.40 (0.33, 0.47)	0.41 (0.34, 0.52)	0.002
AFC	18 (15, 21)	18 (15, 21)	18 (16, 20)	0.98
AMH (μg/L)	6.71 (5.61, 8.42)	6.73 (5.58, 8.43)	6.7 (5.64, 8.4)	0.806
Fertilization, n(%)				0.807
IVF	554 (74.56)	248 (75.15)	306 (74.09)	
ICSI	189 (25.44)	82 (24.85)	107 (25.91)	
Initial dose of Gn	125 (112.5, 150)	125 (112.5, 150)	137.5 (112.5, 150)	< 0.001
Total dose of Gn	2237.5 (1800, 2762.5)	2175 (1825, 2675)	2300 (1775, 2800)	0.535
Dosing days of Gn	12 (11, 14)	12 (11, 14)	12 (11, 13)	0.056
FSH on HCG day (IU/L)	10.62 (8.64, 13.06)	10.3 (8.36, 12.48)	10.84 (8.82, 13.44)	0.018
LH on HCG day (IU/L)	1 (0.8, 1.5)	1 (0.8, 1.5)	1 (0.79, 1.52)	0.787
E2 on HCG day (pg/mL)	3267 (2314.5, 4326)	3213 (2372.25, 4287.25)	3317 (2232, 4332)	0.774
P on HCG day (ng/mL)	0.65 (0.47, 0.91)	0.6 (0.41, 0.85)	0.69 (0.5, 1)	0.003
Endometrial thickness on HCG day (mm)	11.8 (10.5, 13)	12 (10.53, 13)	11.6 (10.3, 13)	0.483
Trigger strategy, n(%)				0.476
rHCG	282 (37.95)	129 (39.09)	153 (37.05)	
GnRH-a combined with HCG	461 (62.05)	201 (60.91)	260 (62.95)	
Number of oocytes retrieved	13 (10, 17)	13 (10, 16.75)	14 (10, 18)	0.228
Number of mature oocytes	8 (5, 11)	8 (5, 12)	8 (4, 11)	0.064
Number of fertilized eggs	10 (7, 13)	10 (7, 13)	10 (7, 13)	0.417
Number of high-quality Cleavage embryos	5 (3, 8)	5 (4, 8)	5 (3, 8)	0.014
Stage of embryos transferred, n(%)				< 0.001
Cleavage	645 (86.81)	245 (74.24)	400 (96.85)	
Blastocyst	98 (13.19)	85 (25.76)	13 (3.15)	
Number of embryos transferred, n(%)				< 0.001
1	245 (32.97)	86 (26.06)	159 (38.5)	
2	498 (67.03)	244 (73.94)	254 (61.5)	

Note: Continuous variables are presented as medians (P25, P75), and categorical data were reported as numbers (%). * $P < 0.05$ was considered statistically significant. BMI, body mass index; FSH, follicle stimulating hormone; LH, luteinizing hormone; E₂, Estradiol; T, testosterone; AMH, anti Müllerian hormone; AFC, antral follicle count; Gn, gonadotropin; hCG, human chorionic gonadotropin; P, progesterone; GnRH-a, gonadotropin-releasing hormone agonist

comprises 7 features, representing the final feature set considered for inclusion in the model.

Development and validation of prediction models

The selected predictive factors were integrated into seven distinct ML models. To optimize the predictive capabilities of each model, hyperparameters underwent further refinement, and 5-fold cross-validation was utilized for performance assessment. The outcomes of these

models on the validation set are presented in Table 3. As depicted in Fig. 3, the XGBoost model achieved superior performance compared to the other six models, attaining an AUC of 0.822 (95% confidence interval [CI] = 0.777–0.867), an accuracy rate of 0.752, specificity of 0.732, sensitivity of 0.772, PPV of 0.682, NPV of 0.812, an F1 score of 0.724, and a Brier score of 0.172 in the validation set. Furthermore, calibration curves were employed to ascertain the predictive effectiveness of the models. The

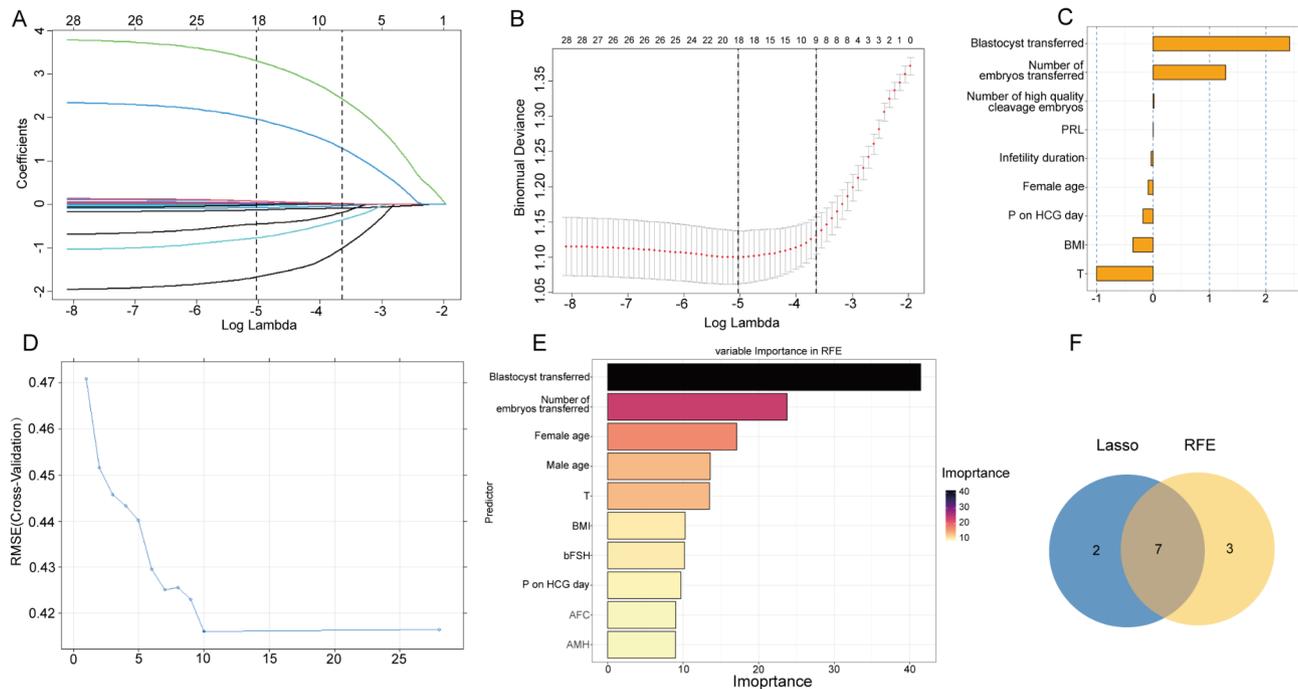


Fig. 2 Features selected by LASSO and RFE. **(A)** The Lasso regression coefficient profiles of all baseline characteristics. **(B)** The optimal lambda selection in the Lasso regression with 10-fold cross-validation. Misclassification errors of different variables against log(lambda) are revealed. The two vertical dashed lines represent the optimal value under the minimum criterion and 1-SE criterion, respectively. The “lambda” is the tuning parameter. **(C)** A total of 9 predictors with non-zero coefficients are identified. **(D)** Features selected by RFE, When the number of features is 10, the RMSE is the lowest. **(E)** The top ten significant predictors identified by RFE. **(F)** The Venn diagram of features selected by LASSO and RFE. The intersection results of two methods yield 7 predictors. LASSO, Least Absolute Shrinkage and Selection Operator; RFE, Recursive Feature Elimination; RMSE, Root Mean Square Error

Table 3 Performance of seven machine learning-based models for predicting live birth in the testing set

Model	AUC	Accuracy	Precision	Sensitivity	Specificity	PPV	NPV	F1 score	Brier score
DT	0.773	0.679	0.669	0.619	0.738	0.669	0.694	0.643	0.194
KNN	0.719	0.643	0.594	0.581	0.705	0.594	0.694	0.587	0.258
LGBM	0.705	0.642	0.605	0.551	0.732	0.605	0.687	0.551	0.215
NBM	0.764	0.720	0.671	0.691	0.749	0.671	0.765	0.577	0.207
RF	0.794	0.702	0.669	0.64	0.765	0.669	0.741	0.654	0.184
SVM	0.806	0.266	0.202	0.243	0.29	0.202	0.34	0.221	0.461
XGB	0.822	0.752	0.682	0.772	0.732	0.682	0.812	0.724	0.172

Note: DT, decision tree; KNN, k-nearest neighbors; LGBM, light gradient boosting machine; NBM, naïve bayes model; RF, random forest; SVM, Support Vector Machine; XGB, eXtreme gradient boosting; AUC, area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value

calibration curve for the validation set in Fig. 4A reflects a high level of alignment between the forecasted live birth rates and the actual outcomes when the XGBoost model is used. Figure 4B displays the decision curve, suggesting that the XGBoost model provides enhanced net benefits when the anticipated live birth rates span from 10 to 90%.

Model interpretability using SHAP

The XGBoost model exhibited excellent predictive capability, leading to the adoption of the SHAP framework for further model interpretation. Figure 5A displays the seven most influential factors, ranked by their mean absolute SHAP values, which include, in descending

order: number of transferred embryos, blastocyst transfer, female age, duration of infertility ≥ 3 , BMI ≥ 28 , testosterone (T) level and P level on HCG day. Figure 5B visualizes the effects of these factors on live birth outcomes, with the y-axis representing the factor values and the x-axis reflecting their influence on the likelihood of live birth. A higher female age, elevated serum T levels, increased P levels on HCG day, duration of infertility ≥ 3 and BMI ≥ 28 were linked with a lower probability of live birth after fresh embryo transfer in PCOS patients. Two representative cases are presented to illustrate personalized feature attributions and demonstrate the application of SHAP in explaining individual model predictions. Specifically, Fig. 5C depicts a PCOS patient who achieved

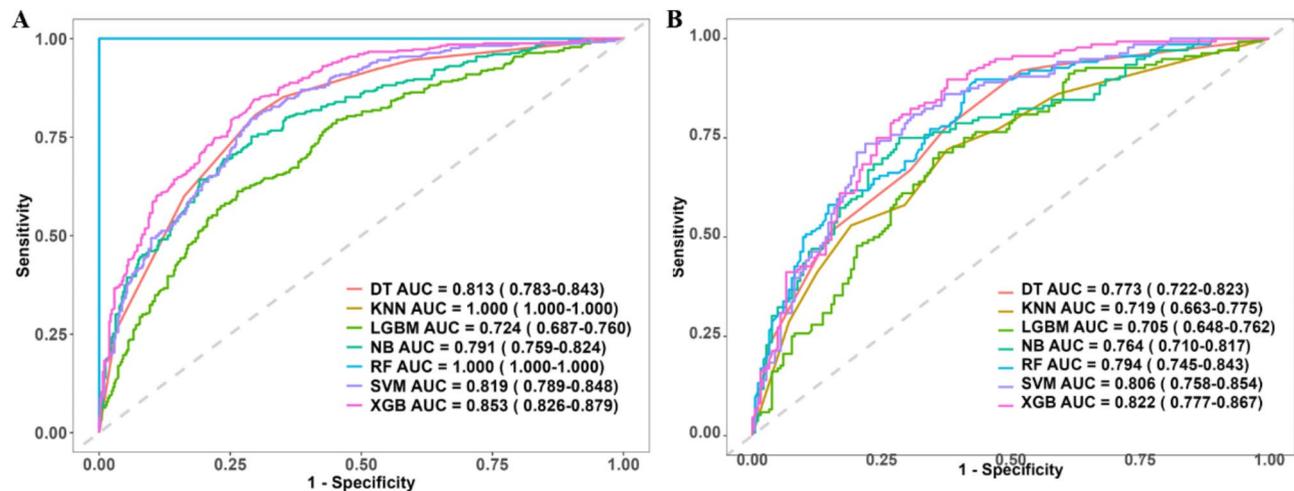


Fig. 3 Comparison of receiver operator characteristic curves (ROCs) for the machine learning models. **(A)** The ROCs of training models. **(B)** The ROCs of validation models. AUC, area under the ROC; DT, decision tree; KNN, k-nearest neighbors; LGBM: light gradient boosting machine; NB, naïve bayes model; RF, random forest; XGBoost, eXtreme gradient boosting

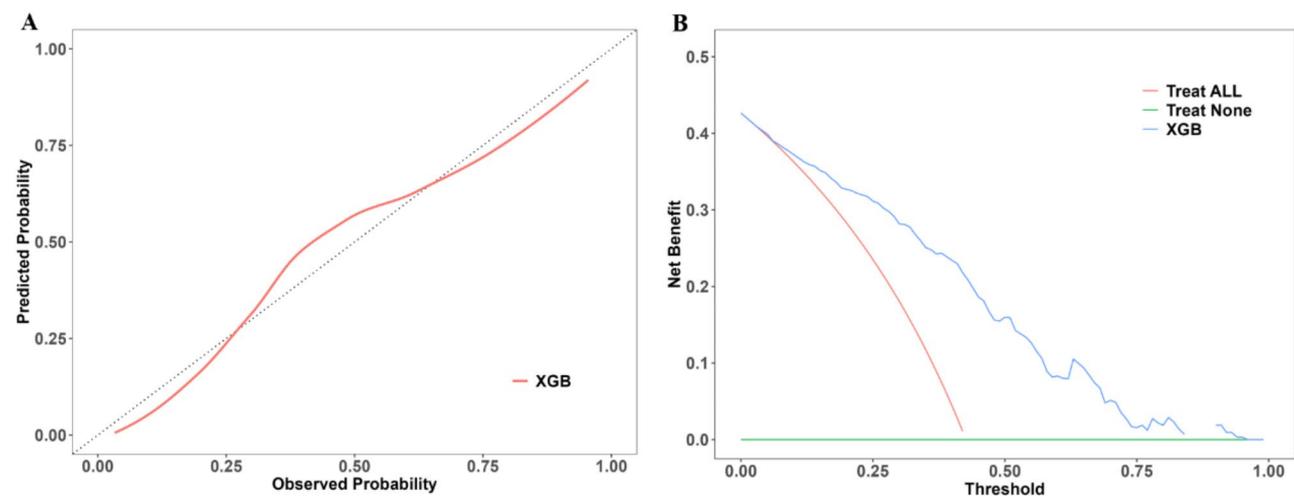


Fig. 4 Discriminative power and accuracy of XGBoost model. **A** The calibration curves of the validation group in XGBoost model. **B** The clinical decision curves of the validation group in XGBoost model

live birth, while Fig. 5D illustrates a PCOS patient who did not achieve live birth. The explanation process begins with the base value, which represents the average prediction across all instances. Subsequently, each input feature at varying levels can either increase or decrease the predicted probability of the outcome. The length of the arrows in the force plots reflects the magnitude of the SHAP values for these features. Ultimately, the model's predicted output for a specific patient is derived.

Discussion

In recent years, the correction of metabolic and endocrine abnormalities, combined with low-dose GN ovarian stimulation in IVE, has enabled a growing number of PCOS patients to undergo fresh embryo transfer, thereby reducing the time required to achieve live birth. Utilizing

ML methods, this study was among the first to explore the determinants affecting live birth outcomes in PCOS patients after fresh embryo transfer. Seven ML prediction models were developed, each exhibiting strong predictive ability in differentiating live birth outcomes within this group, with the XGBoost model delivering the most favorable performance. This model supports clinicians in initiating early diagnostic interventions for these patients and offers valuable insights for enhancing pregnancy outcomes in PCOS cases in the future.

ML techniques exhibit enhanced performance over traditional statistical methods when managing complex relationships among numerous features [15–16], as they can identify influencing factors that might be overlooked by conventional approaches based on experience [17]. For predictor selection, LASSO regression and RFE

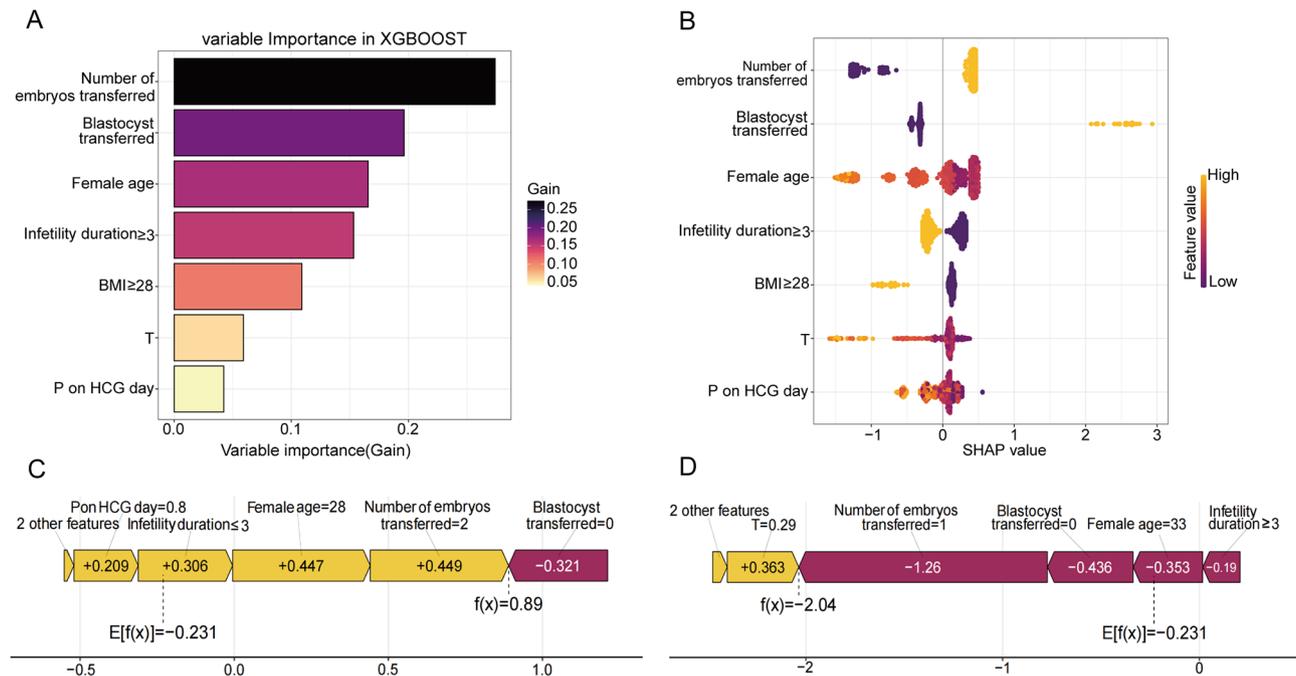


Fig. 5 SHAP plots. **(A)** SHAP summary plot shows feature importance for each predictor of the XGBoost model in descending order. The upper predictors are more important to the model's predictive outcome. A dot is created for each feature attribution value for the XGBoost model of each patient. The further away a dot is from the baseline SHAP value of zero, the stronger it effects the model output. Dots are colored according to the values of features. Yellow represents higher feature values and red represents lower feature values. **(B)** Bar chart of the mean absolute SHAP value for each predictor of the XGBoost model in descending order. **C** and **D**. The force plots provide personalized feature attributions using two examples

were employed, and their intersection was used to construct predictive models. Among the seven ML models developed, the XGBoost model demonstrated the highest performance, achieving an AUC of 0.822 (95% CI=0.777–0.867). To further interpret the model and assess the contributions of individual predictors, SHAP analysis was applied to the top-performing XGBoost model. Each SHAP value quantifies the positive or negative impact of a feature on live birth outcomes after fresh embryo transfer in PCOS patients. Among the predictors, blastocyst transfer provided the most substantial contribution to model predictions, showing higher live birth rates in contrast to cleavage-stage embryo transfer [18]. Nevertheless, considering the heightened possibility of ovarian hyperstimulation syndrome in fresh embryo transfers for PCOS patients, some researchers have proposed adopting a freeze-all strategy [19–20]. A randomized controlled trial involving 1,650 patients compared fresh blastocyst transfer cycles with freeze-all and thawed blastocyst transfer groups, focusing on primary outcomes such as singleton live birth rates and secondary outcomes including pregnancy complications, neonatal birth weight, birth defects, and perinatal complications. The findings revealed that the freeze-all strategy significantly enhanced blastocyst implantation rates, live birth rates, and singleton newborn birth weights, contributing to improved maternal-fetal safety and clinical outcomes.

However, the study also noted that frozen-thawed single blastocyst transfers were linked to an elevated risk of maternal preeclampsia, raising critical considerations for clinical application [21]. In a recent retrospective cohort study of 10,964 single blastocyst transfer cycles, it was observed that transferring single low-grade blastocysts yielded approximately 30% lower live birth rates relative to 44% for high-quality single blastocysts (with very low-grade blastocysts achieving 14%) without negatively affecting perinatal outcomes [22]. SHAP analysis further demonstrated that transferring two embryos increased the probability of live birth. Although twin embryo transfer yields higher pregnancy rates compared to single embryo transfer [23], existing research and consensus emphasize that it also elevates the risk of multiple pregnancies, along with associated pregnancy complications and adverse perinatal outcomes [24–25]. Consequently, single blastocyst transfer is recommended for PCOS patients, as it balances embryo implantation and live birth success with a reduced risk of multiple pregnancies.

As women age, fertility gradually declines. Once women surpass 35 years of age, the likelihood of spontaneous abortion increases substantially, while pregnancy and live birth rates decrease, accompanied by a heightened risk of various pregnancy and perinatal complications [26]. The connection between maternal age and clinical outcomes following embryo transfer in assisted

reproductive technology (ART) has been well-established in numerous studies, with a general consensus highlighting significantly reduced clinical pregnancy and live birth rates in women older than 35 years undergoing ART [27–28]. Consistent findings were observed in this study. Advanced maternal age elevates the risk of early miscarriage, primarily due to a decline in oocyte quality associated with aging, which results in an increased likelihood of embryonic aneuploidy [29]. Preimplantation genetic screening prior to embryo transfer may help mitigate the risk of aneuploidy and reduce the incidence of early miscarriage, although the safety and long-term risks of this approach remain a topic of debate [30]. This study also found that increased progesterone levels on HCG day negatively impacted live birth rates in individuals with PCOS undergoing fresh embryo transfer. Some research suggests that progesterone levels on HCG day may indirectly influence endometrial receptivity and embryo attachment through alterations in gene expression during the implantation window [31]. However, other studies report no association between progesterone levels on HCG day and clinical pregnancy or miscarriage following IVF protocol [32]. As a result, the effect of progesterone measurements on HCG day regarding clinical results post-embryo transfer remains a debated topic.

A high BMI contributes not only to cardiovascular and metabolic disorders but also impairs fertility [33]. Multiple studies have established that a BMI ≥ 28 is correlated with reduced ovarian responsiveness and unfavorable pregnancy outcomes [34–35]. In this investigation, BMI ≥ 28 was ascertained as a prominent risk factor for the failure to achieve live birth in individuals with PCOS undergoing fresh embryo transfer. The foundational treatment for patients with PCOS involves lifestyle interventions, which encompass a diversified approach including appropriate exercise, dietary control, and behavioral modification. For obese PCOS patients, weight loss has been demonstrated to significantly improve treatment outcomes. Specifically, reducing body weight by 5–10% can lead to notable improvements in ovulation, menstrual cycle regulation, and insulin sensitivity. It is important to emphasize that weight loss should be gradual and sustained over time. HA is recognized as a defining feature of PCOS [36]. However, prior investigations into the influence of HA on reproductive outcomes in PCOS patients have been limited, concentrating mainly on early reproductive stages. Research on subsequent maternal and neonatal outcomes in individuals who achieved clinical pregnancy remains scarce, and the existing literature presents some inconsistencies. In animal experiments, Diao et al. [37] found that high-dose androgen exposure could disrupt endometrial development and interfere with the prostaglandin system, potentially causing early pregnancy loss. Similarly, De Vos et al. [38] reported

significantly lower cumulative live birth rates among PCOS patients with HA compared to those without HA. Accordingly, lowering serum testosterone levels in PCOS patients positively influences live birth rates.

Anti-müllerian hormone (AMH) and antral follicle count (AFC) serve as crucial markers for evaluating ovarian reserve function in ART-assisted pregnancy populations. These markers are often utilized as predictors of IVF/ICSI-ET success rates and are regarded as useful metrics for forecasting live birth outcomes. Fertility counseling provided by clinicians frequently relies on changes in these indicators throughout ovarian stimulation to offer personalized guidance [39]. Nevertheless, the findings of this study suggest that AMH and AFC do not exhibit marked predictive value for live birth rates after embryo transfer in PCOS patients, highlighting the necessity for further investigations into the diagnosis and management of PCOS patients undergoing assisted reproductive treatments.

Certain constraints need to be recognized in this investigation. First, the data utilized were obtained from a single center, which may restrict the model's applicability to patients from other institutions. Second, certain parameters, including insulin and glucose metabolism, were absent from the available electronic medical records. Furthermore, external validation of the constructed prediction model has not been performed, raising concerns regarding its generalizability and highlighting the need for further verification. Moving forward, comprehensive external validation datasets will be gathered to enhance the model's robustness.

Seven ML models were developed in this study to predict live birth following fresh embryo transfer in patients with PCOS, demonstrating strong evaluation accuracy. These models provide critical support for identifying high-risk cases within this population that do not result in live birth, facilitate informed treatment decisions, and enable effective monitoring of patient progression.

Author contributions

Yan Sun and Beihong Zheng contributed to the research design. Xiaojing Chen and Wenwen Jiang were responsible for collecting the patient data. Suqin Zhu, Zhiqing Huang, and Yuan Zhou conducted the analysis and interpretation of the patient data. Suqin Zhu and Zhiqing Huang participated in drafting the manuscript. All authors have reviewed and approved the final version of the manuscript.

Funding

This study was supported by grants from the major Scientific Research Program for Young and Middle-aged Health Professionals of Fujian Province (No. 2022ZQNZD010), the Fujian provincial health technology project (No. 2022GGA035), the Fujian Provincial Natural Science Foundation of China (No. 2023J011221, No. 2023J011229 and No. 2020J05279), the Joint Funds for the innovation of science and Technology, Fujian province (No. 2024Y9572, No. 2020Y9137 and No. 2020Y9140), and the innovation Platform Project of Science and Technology, Fujian Province (No. 2021Y2012).

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding authors on reasonable request.

Declarations

Ethics approval and consent to participate

This study was performed in accordance with the Declaration of Helsinki and was approved by the Medical Research Ethics Committee of Fujian Maternity and Child Health Hospital (Ethics approval number:2024KY055). Informed patient consent was not required as the study was retrospective in nature and analyzed patient data anonymously. The requirement for informed consent was waived by the Medical Research Ethics Committee of Fujian Maternity and Child Health Hospital.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Center of Reproductive Medicine, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Maternity and Child Health Hospital, Fujian Medical University, No. 18 Daoshan Road, Fuzhou City 350001, Fujian Province, China

²Fujian Maternal-Fetal Clinical Medicine Research Center, Fuzhou 350001, China

³Fujian Key Laboratory of Prenatal Diagnosis and Birth Defect, Fuzhou 350001, China

Received: 13 November 2024 / Accepted: 26 March 2025

Published online: 04 April 2025

References

1. Qin JZ, Pang LH, Li MJ, Fan XJ, Huang RD, Chen HY. Obstetric complications in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Reprod Biol Endocrinol*. 2013;11:56. <https://doi.org/10.1186/1477-7827-11-56>. Published 2013 Jun 26.
2. Palomba S, Falbo A, Daolio J, Battaglia FA, La Sala GB. Pregnancy complications in infertile patients with polycystic ovary syndrome: updated evidence. *Minerva Ginecol*. 2018;70(6):754–60. <https://doi.org/10.23736/S0026-4784.18.04230-2>.
3. Foroozanfar F, Asemi Z, Bazarganipour F, Taghavi SA, Allan H, Aramesh S. Comparing pregnancy, childbirth, and neonatal outcomes in women with different phenotypes of polycystic ovary syndrome and healthy women: a prospective cohort study. *Gynecol Endocrinol*. 2020;36(1):61–5. <https://doi.org/10.1080/09513590.2019.1631278>.
4. Safiri S, Noori M, Nejadghaderi SA, et al. Prevalence, incidence and years lived with disability due to polycystic ovary syndrome in 204 countries and territories, 1990–2019. *Hum Reprod*. 2022;37(8):1919–31. <https://doi.org/10.1093/humrep/deac091>.
5. Pavlovic ZJ, Jiang VS, Hariton E. Current applications of artificial intelligence in assisted reproductive technologies through the perspective of a patient's journey. *Curr Opin Obstet Gynecol*. 2024;36(4):211–7. <https://doi.org/10.1097/GCO.0000000000000951>.
6. Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Med Inf Decis Mak*. 2019;19(1):211. <https://doi.org/10.1186/s12911-019-0918-5>. Published 2019 Nov 6.
7. Stafford IS, Kellermann M, Mossotto E, Beattie RM, MacArthur BD, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *NPJ Digit Med*. 2020;3:30. <https://doi.org/10.1038/s41746-020-0229-3>. Published 2020 Mar 9.
8. Verma P, Maan P, Gautam R, Arora T. Unveiling the role of artificial intelligence (AI) in polycystic ovary syndrome (PCOS) diagnosis: A comprehensive review. *Reprod Sci*. 2024;31(10):2901–15. <https://doi.org/10.1007/s43032-024-01615-7>.
9. Cheredath A, Uppangala S, C S A, et al. Combining machine learning with metabolomic and embryologic data improves embryo implantation prediction. *Reprod Sci*. 2023;30(3):984–94. <https://doi.org/10.1007/s43032-022-01071-1>.
10. Elmannai H, El-Rashidy N, Mashal I, et al. Polycystic ovary syndrome detection machine learning model based on optimized feature selection and explainable artificial intelligence. *Diagnostics (Basel)*. 2023;13(8):1506. <https://doi.org/10.3390/diagnostics13081506>. Published 2023 Apr 21.
11. Ko YR, Yu EJ, Park JE, et al. Predictive model of IVF outcomes for polycystic ovarian morphology and polycystic ovary syndrome in GnRH antagonist protocol using AMH-MoM and ovarian sensitivity index. *J Obstet Gynaecol Res*. 2024;50(9):1667–74. <https://doi.org/10.1111/jog.16009>.
12. Alenezi SA, Khan R, Amer S. The Impact of High BMI on Pregnancy Outcomes and Complications in Women with PCOS Undergoing IVF-A Systematic Review and Meta-Analysis. *J Clin Med*. 2024;13(6):1578. Published 2024 Mar 10. <https://doi.org/10.3390/jcm13061578>
13. Rotterdam ESHRE, ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril*. 2004;81(1):19–25. <https://doi.org/10.1016/j.fertnstert.2003.10.004>.
14. Song Y, Li R. Interpretation of Chinese guidelines for diagnosis and treatment of polycystic ovary syndrome. *J Pract Obstet Gynecol*. 2018;34(10):737–41. doi: CNKI: SUN: SFCZ.0.2018-10-010.
15. Li D, Kulasegaram K, Hodges BD. Why we needn't fear the machines: opportunities for medicine in a machine learning world. *Acad Med*. 2019;94(5):623–5. <https://doi.org/10.1097/ACM.0000000000002661>.
16. 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. <https://doi.org/10.1016/j.fertnstert.2023.05.149>.
17. Parikh RB, Manz C, Chivers C, et al. Machine learning approaches to predict 6-Month mortality among patients with cancer. *JAMA Netw Open*. 2019;2(10):e1915997. <https://doi.org/10.1001/jamanetworkopen.2019.15997>. Published 2019 Oct 2.
18. Arab S, Badegiesh A, Aldaheri S, Son WY, Dahan MH. What are the live birth and multiple pregnancy rates when 1 versus 2 Low-Quality blastocysts are transferred in A cryopreserved cycle?? A retrospective cohort study, stratified for age, embryo quality, and oocyte donor cycle?. *Reprod Sci*. 2021;28(5):1403–11. <https://doi.org/10.1007/s43032-020-00404-2>.
19. Chen ZJ, Shi Y, Sun Y, Fresh versus Frozen Embryos for Infertility in the Polycystic Ovary Syndrome [published correction appears in N, Engl et al. *J Med*. 2016;375(20):2010. doi: 10.1056/NEJMc160029]. *N Engl J Med*. 2016;375(6):523–533. <https://doi.org/10.1056/NEJMoa1513873>
20. Vuong LN, Dang VQ, Ho TM, et al. IVF transfer of fresh or frozen embryos in women without polycystic ovaries. *N Engl J Med*. 2018;378(2):137–47. <https://doi.org/10.1056/NEJMoa1703768>.
21. Wei D, Liu JY, Sun Y, et al. Frozen versus fresh single blastocyst transfer in ovulatory women: a multicentre, randomised controlled trial. *Lancet*. 2019;393(10178):1310–8. [https://doi.org/10.1016/S0140-6736\(18\)32843-5](https://doi.org/10.1016/S0140-6736(18)32843-5).
22. Zou H, Kemper JM, Hammond ER, et al. Blastocyst quality and reproductive and perinatal outcomes: a multinational multicentre observational study. *Hum Reprod*. 2023;38(12):2391–9. <https://doi.org/10.1093/humrep/dead212>.
23. Attawet J, Wang AY, Farquhar CM, Jordan V, Li Z, Sullivan EA. Pregnancy and birth outcomes of single versus multiple embryo transfer in gestational surrogacy arrangements: a systematic review and meta-analysis. *Hum Fertil (Camb)*. 2022;25(2):217–27. <https://doi.org/10.1080/14647273.2020.1785643>.
24. Wang AY, Safi N, Ali F et al. Neonatal outcomes among twins following assisted reproductive technology: an Australian population-based retrospective cohort study. *BMC Pregnancy Childbirth*. 2018;18(1):320. Published 2018 Aug 8. <https://doi.org/10.1186/s12884-018-1949-0>
25. Ibiebele I, Humphries JB, Torvaldsen S, et al. Gestational age, morbidity and mortality among twin births in new South Wales, Australia 2003–2014: A cohort study. *Aust N Z J Obstet Gynaecol*. 2020;60(4):541–7. <https://doi.org/10.1111/ajo.13101>.
26. Demko ZP, Simon AL, McCoy RC, Petrov DA, Rabinowitz M. Effects of maternal age on euploidy rates in a large cohort of embryos analyzed with 24-chromosome single-nucleotide polymorphism-based preimplantation genetic screening. *Fertil Steril*. 2016;105(5):1307–13. <https://doi.org/10.1016/j.fertnstert.2016.01.025>.
27. Heazell AEP, Newman L, Lean SC, Jones RL. Pregnancy outcome in mothers over the age of 35. *Curr Opin Obstet Gynecol*. 2018;30(6):337–43. <https://doi.org/10.1097/GCO.0000000000000494>.

28. Zhang N, Law YJ, Venetis CA, Chambers GM, Harris K. Female age is associated with the optimal number of oocytes to maximize fresh live birth rates: an analysis of 256,643 fresh ART cycles. *Reprod Biomed Online*. 2021;42(3):669–78. <https://doi.org/10.1016/j.rbmo.2020.11.010>.
29. Soler A, Morales C, Mademont-Soler I, et al. Overview of chromosome abnormalities in first trimester miscarriages: A series of 1,011 consecutive chorionic villi sample karyotypes. *Cytogenet Genome Res*. 2017;152(2):81–9. <https://doi.org/10.1159/000477707>.
30. Yang AM, Xu X, Han Y, et al. Risk factors for different types of pregnancy losses: analysis of 15,210 pregnancies after embryo transfer. *Front Endocrinol (Lausanne)*. 2021;12:683236. <https://doi.org/10.3389/fendo.2021.683236>. Published 2021 Jun 25.
31. Yang Y, Liu B, Wu G, Yang J. Exploration of the value of progesterone and progesterone/estradiol ratio on the hCG trigger day in predicting pregnancy outcomes of PCOS patients undergoing IVF/ICSI: a retrospective cohort study. *Reprod Biol Endocrinol*. 2021;19(1):184. <https://doi.org/10.1186/s12958-021-00862-6>. Published 2021 Dec 11.
32. Saharkhiz N, Salehpour S, Tavasoli M, Aghighi A. Premature progesterone rise at human chorionic gonadotropin triggering day has no correlation with intracytoplasmic sperm injection outcome. *Iran J Reprod Med*. 2015;13(2):79–84.
33. Practice Committee of the American Society for Reproductive Medicine. Electronic address: Asrm@asrm.org; practice committee of the American society for reproductive medicine. Obesity and reproduction: a committee opinion. *Fertil Steril*. 2021;116(5):1266–85. <https://doi.org/10.1016/j.fertnstert.2021.08.018>.
34. Broughton DE, Moley KH. Obesity and female infertility: potential mediators of obesity's impact. *Fertil Steril*. 2017;107(4):840–7. <https://doi.org/10.1016/j.fertnstert.2017.01.017>.
35. Romanski PA, Bortoletto P, Magaoay B, Chung A, Rosenwaks Z, Spandorfer SD. Live birth outcomes in infertile patients with class III and class IV obesity following fresh embryo transfer. *J Assist Reprod Genet*. 2021;38(2):347–55. <https://doi.org/10.1007/s10815-020-02011-1>.
36. Teede HJ, Tay CT, Laven JJE, et al. Recommendations from the 2023 international Evidence-based guideline for the assessment and management of polycystic ovary syndrome. *J Clin Endocrinol Metab*. 2023;108(10):2447–69. <https://doi.org/10.1210/clinem/dgad463>.
37. Diao HL, Su RW, Tan HN, et al. Effects of androgen on embryo implantation in the mouse delayed-implantation model. *Fertil Steril*. 2008;90(4 Suppl):1376–83. <https://doi.org/10.1016/j.fertnstert.2007.07.1341>.
38. De Vos M, Pareyn S, Drakopoulos P, et al. Cumulative live birth rates after IVF in patients with polycystic ovaries: phenotype matters. *Reprod Biomed Online*. 2018;37(2):163–71. <https://doi.org/10.1016/j.rbmo.2018.05.003>.
39. Metello JL, Tomás C, Ferreira P. Can we predict the IVF/ICSI live birth rate? *JBRA Assist Reprod*. 2019;23(4):402–7. <https://doi.org/10.5935/1518-0557.20190043>. Published 2019 Oct 14.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.