Reproductive Biology and Endocrinology

RESEARCH Open Access

Identifying key predictive features for live birth rate in advanced maternal age patients undergoing single vitrified-warmed blastocyst transfer

Lidan Liu[1*†], Bo Liu[1†], Ming Liao[1], Qiuying Gan[2], Qianyi Huang[1] and Yihua Yang[1*]

Abstract

Background Infertility affects one in six couples worldwide, with advanced maternal age (AMA) posing unique challenges due to diminished ovarian reserve and reduced oocyte quality. Single vitrified-warmed blastocyst transfer (SVBT) has shown promise in assisted reproductive technology (ART), but success rates in AMA patients remain suboptimal. This study aimed to identify and refine predictive factors for live birth following SVBT in AMA patients, with the goal of enhancing clinical decision-making and enabling personalized treatment strategies.

Methods This retrospective cohort study analyzed 1,168 SVBT cycles conducted between June 2016 and December 2022 at the First Affiliated Hospital of Guangxi Medical University and Nanning Maternity and Child Health Hospital. Nineteen machine-learning models were applied to identify key predictive factors for live birth. Feature selection and 10-fold cross-validation were employed to validate the models.

Results The most significant predictors of live birth included inner cell mass quality, trophectoderm quality, number of oocytes retrieved, endometrial thickness, and the presence of 8-cell blastomeres on day 3. The stacking model demonstrated the best predictive performance (AUC: 0.791), followed by Extra Trees (AUC: 0.784) and Random Forest (AUC: 0.768). These models outperformed traditional methods, achieving superior accuracy, sensitivity, and specificity.

Conclusion Leveraging advanced machine-learning models and identifying critical predictive factors can improve the accuracy of live birth outcome predictions for AMA patients undergoing SVBT. These findings offer valuable insights for enhancing clinical decision-making and managing patient expectations. Further research is needed to validate these results in larger, multi-center cohorts and to explore additional factors, including fresh embryo transfers, to broaden the applicability of these models in clinical practice.

†Lidan Liu and Bo Liu contributed equally to this work.

*Correspondence:
Lidan Liu
liulidan2022@126.com
Yihua Yang
workyyh@163.com

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

Keywords Assisted Reproductive Technology (ART), Advanced maternal age (AMA), Live birth rate, Machine-learning, Predictive modeling, Single vitrified-warmed blastocyst transfer (SVBT)

Introduction

Infertility is a significant global health issue, affecting approximately one in six couples worldwide [1]. Assisted reproductive technology (ART) offers hope, and recent advancements in prolonged embryo culture and vitrification have significantly improved cumulative live birth rates (cLBR) [2]. Extending embryo development to 5–6 days via blastocyst culture enhances embryo selection and uterine synchronization, which boosts implantation rates [3]. Vitrification, a rapid cryopreservation method, prevents ice crystal formation, resulting in higher success rates for frozen embryo transfers (FET) [4]. These innovations have led to the development of segmented IVF cycles, an approach where ovarian stimulation and egg retrieval are conducted in one cycle, but embryo transfer is delayed to a later cycle. In this method, embryos are vitrified (frozen) and transferred when the patient's hormonal conditions are more optimal, improving implantation and pregnancy outcomes. When combined with single vitrified-warmed blastocyst transfer (SVBT), where only one embryo is transferred at a time, this approach further reduces the risk of multiple pregnancies and enhances overall ART success [5, 6].

However, women of advanced maternal age (AMA), defined as 35 years and older, face unique reproductive challenges. Their lower ovarian reserve, reduced oocyte quality, and increased risk of chromosomal abnormalities result in lower ART success rates [6]. Despite progress in SVBT, AMA patients still have suboptimal outcomes compared to younger women, which highlights the need for optimized protocols [7]. Given the financial and emotional burden of IVF, understanding the factors that influence live birth rates in AMA patients is critical for counseling and managing expectations [8].

Traditionally, IVF predictions were based on clinicians' subjective assessments, relying primarily on patient age and success rates of fertility centers [9]. While useful, this method struggles with the complexity of individual cases [10]. Machine learning (ML) offers a solution, automating predictions by analyzing large datasets to identify patterns that traditional methods might miss [11, 12].

ML techniques such as support vector machines (SVM), decision trees (DT), random forests (RF), and extra trees are commonly used for IVF outcome predictions [13]. These models help identify critical factors influencing IVF success by analyzing patient records, offering insights that manual assessments may overlook [12]. These models help identify critical factors influencing IVF success by analyzing patient records, offering insights that manual assessments may overlook [12, 13]. Several studies have used ML to predict IVF outcomes based on clinical variables. Assessing live birth rate using these parameters, available before the IVF cycle, could improve ART efficiency and reduce resource wastage (Li et al., 2023; Qiu et al., 2019; Wang et al., 2022). A 2023 study by Liu et al. developed and validated a model to predict live birth rates based on embryo morphology and vitrification day in SVBT cycles. The results indicated that female age, embryo quality, and vitrification day are closely linked to live birth rates. However, the study did not provide detailed age stratification, potentially affecting the model's practical accuracy. Additionally, the single-center nature of the research limits the generalizability of the findings, and the model's predictive ability remains moderately accurate, with AUC values of 0.66 (training) and 0.65 (validation) [9].

Despite the growing body of research, there is still a lack of predictive models specifically tailored for AMA patients undergoing SVBT. An automated model that analyzes clinical and laboratory data could help clinicians make informed decisions and personalize treatment plans for these patients, potentially improving success rates and reducing the emotional and financial burdens associated with IVF [14].

Therefore, this study aims to identify and evaluate key predictive factors for live birth rate in AMA patients following SVBT. By employing advanced machine learning models, this research seeks to optimize these predictive features to improve ART success rates for AMA patients, ultimately providing more effective treatment strategies for clinicians.

Methods

Study design and participants

This retrospective cohort study was performed at two reproductive medicine centers: the First Affiliated Hospital of Guangxi Medical University and the Nanning Maternity and Child Health Hospital. It analyzed 1,168 single frozen-thawed blastocyst transfer cycles between June 2016 and December 2022. ART treatments were recorded in the ART database following the Technical Standard for Human-Assisted Reproduction by the Chinese Ministry of Health. The study included infertile women aged 35 years or older who underwent single frozen-thawed blastocyst transfers, with no preimplantation genetic testing.

Ethical approval and compliance

The study protocol obtained ethical approval from the Institutional Review Boards of the participating hospitals, specifically the First Affiliated Hospital of Guangxi Medical

University Medical Ethics Committee (Approval Number: 2024-E445-01) and the Nanning Maternity and Child Health Hospital Medical Ethics Committee (Approval Number: YX20240627-1). Informed consent was secured from all participants.The authors confirm that the guidelines approved by the local institution were followed.

Ovarian stimulation and oocyte insemination

Ovarian stimulation protocols were not restricted, allowing for individualization based on the patient's characteristics. The initial dose of recombinant follicle-stimulating hormone (rFSH, Gonal-F, Merck or Puregon, Organon) was determined by age, BMI, baseline FSH levels, and antral follicle counts [15]. Human chorionic gonadotropin (HCG, Ovitrelle, Merck) was administered when at least one follicle reached 18 mm or larger. Oocyte retrieval was performed 36 h post-HCG administration via transvaginal ultrasound-guided aspiration. Fertilization was carried out using either conventional IVF or intracytoplasmic sperm injection (ICSI) based on semen quality, following routine center protocols.

Embryo culture and blastocyst scoring

Embryos were cultured in G-TL media (Vitrolife) from fertilization to the blastocyst stage. The embryos were incubated at 37 °C in an atmosphere containing 5% O_2, 6% CO_2, and nitrogen as the balance gas, under oil. Blastocyst assessment was based on the Gardner grading system [16], evaluating expansion, inner cell mass, and trophectoderm quality.

Blastocyst vitrification and thawing procedures

Fully expanded blastocysts were artificially collapsed using a laser prior to cryopreservation. Vitrification was performed using Cryotop Safety Kits (KITAZATO), with embryos loaded onto cryotops on day 6 post-insemination and stored in liquid nitrogen. For warming, Blastocyst Warming Kits (KITAZATO) were used once the endometrium reached the required thickness for transfer. Blastocyst survival was assessed based on re-expansion two hours post-warming.

Endometrial Preparation and blastocyst transfer

Endometrial preparation for frozen embryo transfer (FET) was conducted using various protocols, including modified natural cycles, mild stimulation cycles, and hormone replacement therapy (HRT) cycles, with or without GnRH agonist pretreatment, as detailed below:

Modified natural cycle protocol

Follicular development was monitored via ultrasound starting on days 10–12 of the menstrual cycle. Once the dominant follicle reached a size of ≥18 mm, ovulation was triggered using recombinant human chorionic gonadotropin (Ovitrelle, Merck). Luteal support commenced after confirming ovulation, with a regimen of oral dydrogesterone (Duphaston, Abbott) 20 mg daily and vaginal progesterone gel (Crinone, Merck) 90 mg daily, facilitating the transition of the endometrium from the proliferative to the secretory phase. Blastocyst transfer was performed five days after ovulation.

Mild stimulation cycle protocal

In this protocol, if the dominant follicle measured < 12 mm by days 10–12, intramuscular human menopausal gonadotropin (LeBaode, Livzon) was administered to promote follicular growth. Ovulation was triggered with recombinant human chorionic gonadotropin (Ovitrelle, Merck) once the follicle reached ≥ 18 mm. Luteal support was initiated using the same regimen as in the modified natural cycle, with blastocyst transfer performed five days post-ovulation.

Hormone replacement therapy (HRT) protocol

HRT was initiated on days 3–5 of the menstrual cycle or after withdrawal bleeding, with estradiol valerate (Progynova, Bayer) administered at 2 mg three times daily. To reduce the risk of thrombosis, oral aspirin (50–100 mg daily) was prescribed, provided no contraindications were present. After 12–15 days, ultrasound was used to assess endometrial thickness, alongside serum estradiol (E2) and progesterone levels. Two options for progesterone support were available: (1) intramuscular progesterone (60 mg/day) combined with oral dydrogesterone (30 mg/day), or (2) vaginal progesterone gel (Crinone, Merck) (90 mg/day) combined with oral dydrogesterone (30 mg/day). The choice of regimen was based on patient preference. Blastocyst transfer was performed on the sixth day after initiating progesterone.

GnRH agonist combined with HRT protocol

This protocol involved downregulation with subcutaneous administration of triptorelin acetate (Diphereline, Ipsen) at 3.75 mg on days 2–5 of menstruation. Hormonal and endometrial parameters were monitored between days 28–35. Downregulation was confirmed when E2 levels were < 50 pg/mL, FSH < 5 IU/L, LH < 5 IU/L, and endometrial thickness < 5 mm. Once these criteria were met, estradiol valerate was administered, and the remaining steps followed the HRT protocol.

Clinical outcomes

The primary outcome measured was live birth, defined as the delivery of any viable infant at 28 weeks of gestation or later. Twins delivered by one mother were considered a single live birth.

Data Collection and candidate predictors

The original dataset included over 40 variables. Based on clinical expert recommendations, 19 variables relevant to live birth were selected: Maternal age at FET, Paternal age at FET, Maternal BMI, Basal FSH, Basal LH, Infertility duration, E2 on trigger day, Total gonadotropin dose, Number of oocytes retrieved, Endometrial thickness, Blastulation time, Blastocyst stage, Inner cell mass, Trophectoderm, Fragmentation on day 3, 8 blastomere on day 3, Infertility type, Fertilization method, and Endometrial preparation.

Data pre-processing and balancing

The study included 1,168 cycles, comprising 352 live birth cycles and 816 non-live birth cycles, with no missing values. Numerical data were standardized using z-score normalization for both training and validation sets to ensure comparability across features. Categorical variables were label-encoded for compatibility with machine learning models.Class imbalance was identified, with the live birth cycles representing approximately 30% of the total dataset. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied specifically to the live birth outcome variable in the training set to balance the data. The risk of potential oversampling and its implications, such as overfitting, were considered, and appropriate precautions were taken, as discussed in Alkhawaldeh et al. (2023). The balanced dataset was then split into training and validation sets at a ratio of 0.75 to 0.25.

The study included 1,168 cycles, comprising 352 live birth cycles and 816 non-live birth cycles, with no missing values. Numerical data were standardized for both training and validation sets. Categorical variables were label-encoded for compatibility with machine-learning models. Given the class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was used to balance the data. The balanced data were divided into training and validation sets at a ratio of 0.75 to 0.25.

Feature selection

Feature selection aimed to enhance model predictability, interpretability, and performance by reducing dataset dimensionality and training time. Four classifiers—random forest, extreme gradient boosting (XGBoost), lasso regression, and extremely randomized trees (Extra Trees)—were applied to identify key predictors for live birth rate. The top four variables from each classifier were combined into a single feature subset, improving the model's generalization on unknown data and ensuring coverage of all critical features.

Machine-learning Approach and evaluation

Various machine-learning models were employed to predict live birth rate, including XGBoost, logistic regression, support vector machine (SVM), random forest, multilayer perceptron (MLP), K-nearest neighbors (KNN), Extra Trees, light gradient boosting machine (LightGBM), gradient boosting, AdaBoost, Bagging, Gaussian Naive Bayes (Gaussian NB), Bernoulli Naive Bayes (Bernoulli NB), decision tree, quadratic discriminant analysis (QDA), ridge classifier, passive aggressive classifier, and CatBoost. The two best-performing models were combined into an ensemble model called Stacked Generalization (stacking) to enhance prediction performance and generalization.

A 10-fold cross-validation method, repeated three times, was used to create training and validation sets. The dataset was initially divided into 10 subsets. In each iteration, one subset served as the validation set while the remaining subsets were used for training. This process was repeated three times with different dataset partitions. Final results were obtained by averaging the outcomes from all rounds, ensuring stability and reliability.Predictive performance was evaluated using receiver operating characteristic (ROC) curves, area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) with 95% confidence intervals (CI).

Statistical analysis

Statistical analysis was performed using Python software (version 3.12). Participant characteristics were summarized with means and standard deviations for continuous variables, and frequencies and percentages for categorical variables. T-tests were used to compare continuous variables, and chi-square tests or Fisher's exact tests were used for categorical variables. A p-value of less than 0.05 was considered statistically significant.

Results

Baseline characteristics of SVBT cycles stratified by live birth rate

This study analyzed a total of 1,168 SVBT cycles involving AMA patients. Among the patients, 30.1% (352/1,168) had a live birth, while 69.9% (816/1,168) did not. Based on clinical expert's opinions, we initially identified 19 variables for live birth rate. Table 1 presents the baseline characteristics of 1,168 SVBT cycles, stratified by live birth. The key findings are as follows: live birth was associated with younger maternal ($p=0.000$) and paternal age ($p=0.001$), a lower gonadotropin dose ($p=0.032$), a higher number of oocytes retrieved ($p=0.003$), greater endometrial thickness ($p=0.015$), and better blastocyst quality (inner cell mass ($p=0.000$) and trophectoderm ($p=0.000$)). Modified natural cycles were more common in the live birth group, while HRT was more frequent in those without live births ($p=0.030$). Blastocyst stage also differed significantly between groups ($p=0.043$). Other variables showed no significant associations with live birth outcomes.

Table 1 Baseline characteristics of 1168 SVBT cycles stratified by live birth outcome

Variable	Live birth ($n=352$)	No live birth ($n=816$)	p value
Maternal age at FET	37.40 ± 2.28	38.20 ± 2.73	0.000***
Paternal age at FET	38.51 ± 4.34	39.41 ± 4.29	0.001**
Maternal BMI	22.13 ± 2.78	22.19 ± 2.64	0.729
Basal FSH	6.51 ± 1.50	6.69 ± 1.67	0.068
Basal LH	5.65 ± 2.56	5.60 ± 2.78	0.772
Infertility duration	5.39 ± 3.91	5.26 ± 4.07	0.627
E_2 on trigger day	4426.04 ± 2335.40	4154.40 ± 2353.84	0.070
Total gonadotropin dose	2458.60 ± 860.21	2575.60 ± 854.31	0.032*
Number of oocyte retrieved	18.90 ± 7.43	17.39 ± 8.00	0.003**
Endometrial thickness	9.75 ± 1.53	9.49 ± 1.72	0.015*
Blastulation time			
Day 5	209 (59.38%)	437 (53.56%)	0.076
Day 6	143 (40.63%)	379 (46.45%)	
Blastocyst stage			0.043*
3	64 (18.18%)	211 (25.86%)	
4	217 61.65%)	461 (56.50%)	
5	65 (18.47%)	131 (16.05%)	
6	6 (1.70%)	13 (1.59%)	
Inner cell mass			0.000***
A	124 (35.23%)	212 (25.98%)	
B	192 (54.55%)	440 (53.92%)	
C	36 (10.23%)	164 (20.10%)	
Trophectoderm			0.000***
A	134 (38.07%)	211 (25.86%)	
B	188 (53.41%)	451 (55.27%)	
C	30 (8.52%)	154 (18.87%)	
Fragmentation on day 3			
≤ 10%	162 (46.02%)	368 (45.10%)	0.634
11-25%	173 (49.15%)	397 (48.65%)	
26-50%	17 (4.83%)	51 (6.25%)	
8 blastomere on day 3			
YES	146 (41.48%)	294 (36.03%)	0.090
NO	206 (58.52%)	522 (63.97%)	
Infertility type			
PI	94 (26.70%)	216 (26.47%)	0.991
SI	258 (73.30%)	600 (73.53%)	
Fertilization method			
IVF	279 (79.26%)	158 (80.64%)	0.644
ICSI	73 (20.74%)	658 (19.36%)	
Endometrial preparation			0.030*
modified natural cycles	144 (40.91%)	267 (32.72%)	
mild stimulation	17 (4.83%)	31 (3.80%)	
HRT	121 (34.38%)	338 (41.42%)	
GnRHa_HRT	70 (19.89%)	180 (22.06%)	

Note: Asterisks indicate levels of statistical significance: * indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$

Screening predictive factors

A total of 1,168 SVBT cycles were divided into training (0.75) and validation (0.25) sets. We used four statistical methods—Random Forest, Lasso Regression, XGBoost, and Extra Trees—to identify predictive factors for live birth rate. Figure 1 ranks the importance of each variable based on its contribution to the model's performance. The top predictive factors included maternal age at FET, inner cell mass (ICM), trophectoderm quality, total gonadotropin dose, 8 blastomeres on day 3 (indicating embryo quality), endometrial preparation, paternal age at FET, number of oocytes retrieved, endometrial thickness, infertility duration, and blastocyst stage. These ten

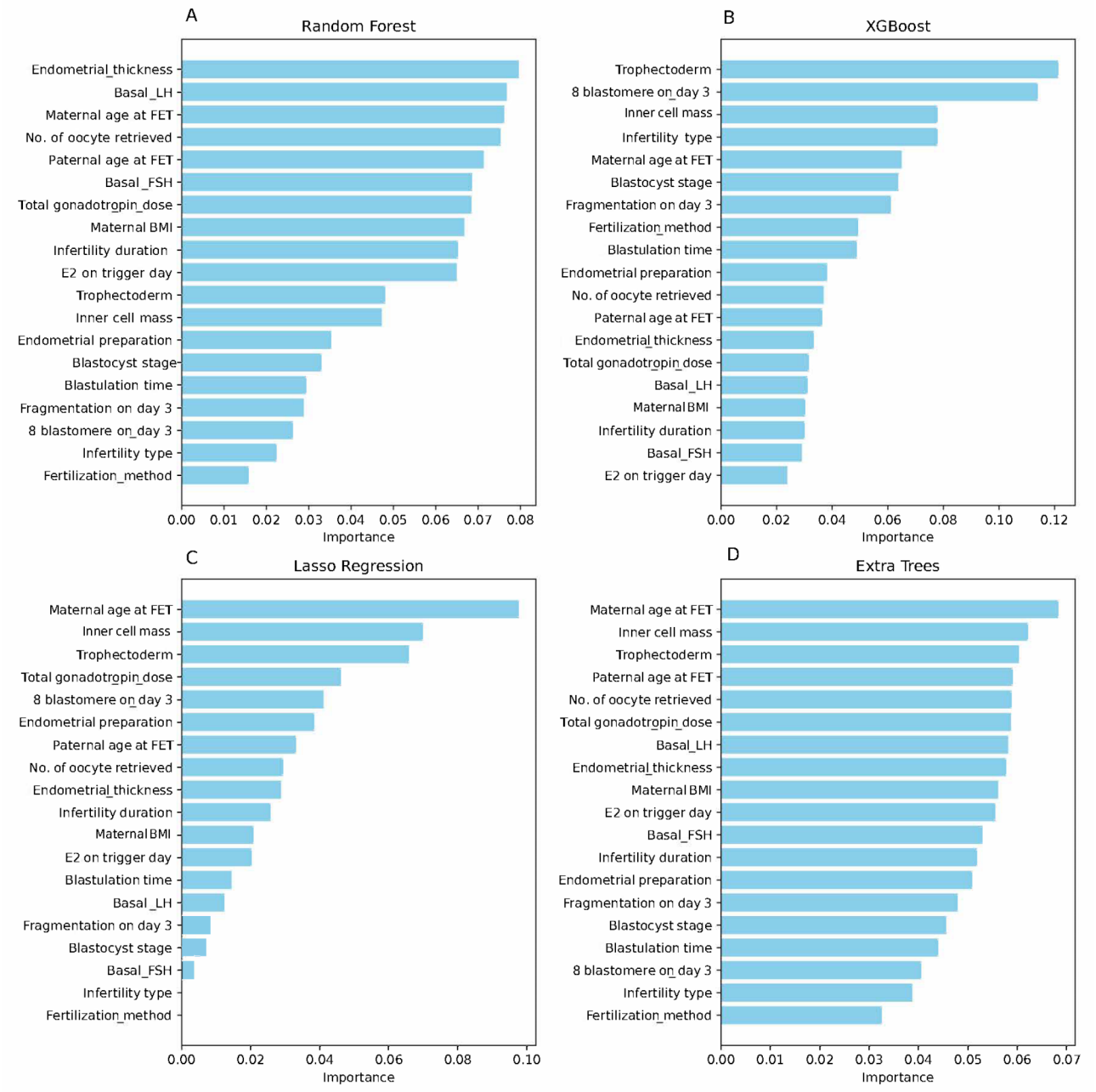

Fig. 1 Ranking of variable importance for four classifiers on the training set: (**A**) Random Forest; (**B**) XGBoost; (**C**) Lasso Regression; (**D**) Extra Tree

factors were selected for further validation and model evaluation.

Comparison of AUC on the Training Set

We compared the AUC values of different machine learning models using ROC curves to evaluate model performance on the training set. A higher AUC indicates better generalization ability. Figure 2 shows the mean AUC values from ten-fold cross-validation for 19 models. The top-performing models were the stacking model (AUC: 0.853±0.034), Extra Trees (AUC: 0.849±0.035), and Random Forest (AUC: 0.832±0.030). Other notable models included XGBoost (AUC: 0.789±0.045) and Gradient Boosting (AUC: 0.801±0.042). The stacking model outperformed all other models.

Performance comparison of various machine-learning models on the Validation Set

Table 2 summarizes the performance of 19 machine-learning models using 10-fold cross-validation on the

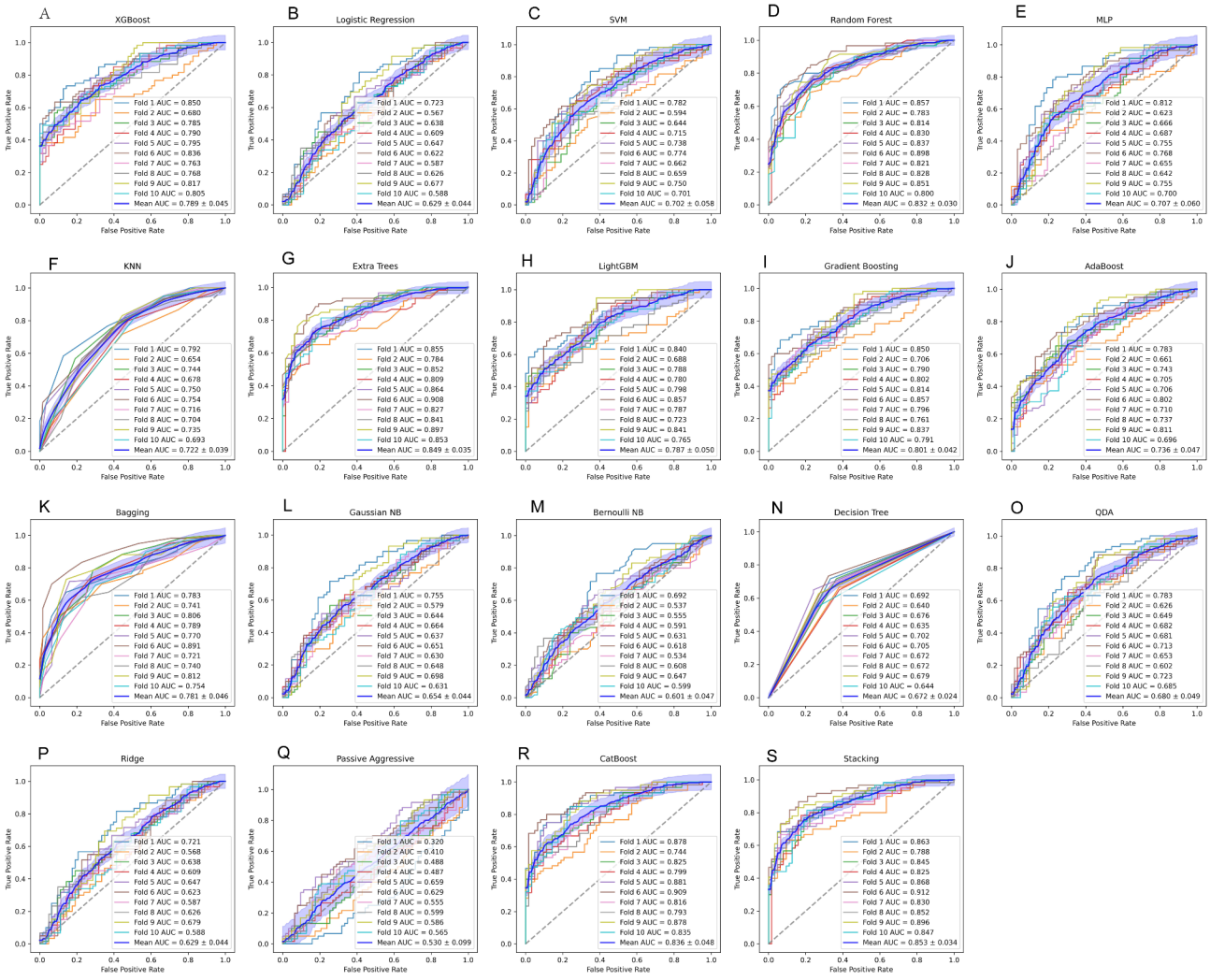

Fig. 2 Comparision of the AUC performance of 19 machine learning models with 10-fold cross-validation on the training set: (**A**) XGBoost model; (**B**) Logistic Regression model; (**C**) SVM model; (**D**) Random Forest model; (**E**) MLP model; (**F**) KNN model; (**G**) Extra Tree model; (**H**) LightGBM model; (**I**) Gradient Boosting model; (**J**) AdaBoost model; (**K**) Bagging model; (**L**) Gaussian NB model; (**M**) Bernoulli NB model; (**N**) Decision Tree model; (**O**) QDA model; (**P**) Ridge model; (**Q**) Passive Aggressive model; (**R**) CatBoost model; (**S**) Stacking model

validation set, evaluating metrics such as accuracy, sensitivity, specificity, precision, NPV, F1 score, and AUC. The Stacking model (AUC: 0.791), Extra Trees (AUC: 0.784), and Random Forest (AUC: 0.768) were the top performers. Other notable models included CatBoost (AUC: 0.768), Gradient Boosting (AUC: 0.732), and XGBoost (AUC: 0.730). Lower-performing models were Passive Aggressive (AUC: 0.561), Bernoulli NB (AUC: 0.554), and Logistic Regression (AUC: 0.583). In summary, the Stacking, Extra Trees, and Random Forest models demonstrated the best predictive performance.

Ranking of feature importance in Predicting Live Birth Rate

Figure 3 ranks the feature importance for predicting live birth rate in AMA patients, using the Mean Decrease Accuracy (MDA) metric from the Extra Trees model. The most important features were inner cell mass (MDA: 0.0569), trophectoderm (MDA: 0.0474), number of oocytes retrieved, and endometrial thickness (both MDA: 0.0466), followed by 8-blastomere on day 3 (MDA: 0.0436). Other moderately important features included basal LH, paternal and maternal age at FET, and total gonadotropin dose. Infertility type had the lowest importance (MDA: 0.0209). The top predictors were inner cell mass, trophectoderm, number of oocytes retrieved, endometrial thickness, and 8-blastomere on day 3.

Discussion

Recent advances in machine learning for IVF outcomes

In the last five years, machine learning (ML) has revolutionized various aspects of Assisted Reproductive Technology (ART), including embryo selection, IVF outcome prediction, and genetic screening. Deep learning models have significantly enhanced the accuracy and consistency

Table 2 Performance comparison of various machine-learning models on the Validation Set

Machine Learning Model	AUC	Accuracy	Sensitivity	Specificity	Precision	NPV	F1
XGBoost	0.730 (0.687–0.774)	0.728 (0.684–0.770)	0.706 (0.644–0.768)	0.753 (0.695–0.819)	0.766 (0.705–0.826)	0.691 (0.628–0.753)	0.735 (0.683–0.782)
Logistic Regression	0.583 (0.534–0.630)	0.583 (0.534–0.627)	0.587 (0.517–0.652)	0.579 (0.511–0.648)	0.615 (0.548–0.682)	0.550 (0.483–0.618)	0.601 (0.542–0.654)
SVM	0.664 (0.621–0.710)	0.669 (0.625–0.713)	0.734 (0.678–0.793)	0.595 (0.524–0.659)	0.675 (0.616–0.733)	0.661 (0.587–0.729)	0.703 (0.654–0.752)
Random Forest	0.768 (0.726–0.806)	0.765 (0.723–0.806)	0.720 (0.662–0.771)	0.816 (0.755–0.871)	0.818 (0.761–0.872)	0.718 (0.659–0.776)	0.766 (0.715–0.808)
MLP	0.677 (0.628–0.720)	0.679 (0.632–0.723)	0.711 (0.652–0.770)	0.642 (0.571–0.708)	0.695 (0.631–0.752)	0.659 (0.587–0.726)	0.703 (0.651–0.749)
KNN	0.663 (0.615–0.708)	0.669 (0.620–0.713)	0.757 (0.697–0.810)	0.568 (0.497–0.640)	0.668 (0.608–0.730)	0.671 (0.602–0.748)	0.710 (0.658–0.756)
Extra Trees	0.784 (0.742–0.818)	0.782 (0.740–0.821)	0.757 (0.698–0.816)	0.811 (0.755–0.865)	0.821 (0.764–0.872)	0.744 (0.689–0.801)	0.788 (0.743–0.829)
Light GBM	0.726 (0.685–0.765)	0.723 (0.679–0.762)	0.683 (0.623–0.750)	0.768 (0.705–0.827)	0.772 (0.713–0.830)	0.679 (0.619–0.737)	0.725 (0.675–0.771)
Gradient Boosting	0.732 (0.688–0.775)	0.730 (0.689–0.777)	0.711 (0.646–0.768)	0.753 (0.695–0.810)	0.767 (0.708–0.824)	0.694 (0.629–0.755)	0.738 (0.694–0.784)
AdaBoost	0.700 (0.652–0.742)	0.701 (0.657–0.748)	0.716 (0.654–0.771)	0.684 (0.620–0.747)	0.722 (0.660–0.784)	0.677 (0.609–0.741)	0.719 (0.667–0.762)
Bagging	0.740 (0.699–0.781)	0.733 (0.689–0.775)	0.638 (0.571–0.702)	0.842 (0.788–0.893)	0.822 (0.761–0.879)	0.669 (0.610–0.724)	0.718 (0.667–0.766)
Gaussian NB	0.617 (0.572–0.665)	0.623 (0.574–0.669)	0.697 (0.638–0.756)	0.537 (0.467–0.608)	0.633 (0.575–0.692)	0.607 (0.532–0.676)	0.664 (0.610–0.706)
Bernoulli NB	0.554 (0.506–0.602)	0.554 (0.505-0.600)	0.550 (0.481–0.616)	0.558 (0.489–0.625)	0.588 (0.517–0.656)	0.520 (0.451–0.587)	0.569 (0.513–0.624)
Decision Tree	0.640 (0.592–0.691)	0.642 (0.598–0.686)	0.665 (0.597–0.724)	0.616 (0.543–0.685)	0.665 (0.602–0.726)	0.616 (0.547–0.687)	0.665 (0.615–0.718)
QDA	0.648 (0.603–0.694)	0.652 (0.608–0.699)	0.706 (0.648–0.763)	0.589 (0.514–0.663)	0.664 (0.602–0.725)	0.636 (0.566–0.714)	0.684 (0.635–0.730)
Ridge	0.588 (0.540–0.634)	0.588 (0.542–0.640)	0.596 (0.534–0.659)	0.579 (0.510–0.644)	0.619 (0.550–0.684)	0.556 (0.483–0.624)	0.607 (0.556–0.661)
Passive Aggressive	0.561 (0.520–0.600)	0.542 (0.495–0.591)	0.284 (0.219–0.346)	0.837 (0.781–0.885)	0.667 (0.573–0.764)	0.505 (0.452–0.564)	0.399 (0.333–0.467)
CatBoost	0.768 (0.726–0.808)	0.765 (0.723–0.806)	0.716 (0.657–0.779)	0.821 (0.764–0.869)	0.821 (0.768–0.872)	0.716 (0.657–0.776)	0.765 (0.718–0.810)
Stacking	0.791 (0.752–0.831)	0.789 (0.748–0.826)	0.761 (0.701–0.815)	0.821 (0.761–0.872)	0.830 (0.776–0.879)	0.750 (0.687–0.806)	0.794 (0.751–0.835)

XG Boost: Extreme Gradient Boosting; SVM, Support Vector Machine; MLP: Multilayer Perceptron; KNN: K-Nearest neighbors; Extra Trees: extremely randomized trees; SVC: support Vector Classifier; light GBM: light gradient boosting machine; GaussianNB: Gaussian Naive Bayes; Bernoulli NB: Bernoulli Naive Bayes; QDA: quadratic discriminant analysis; stacking: stacked generalization

of embryo scoring by analyzing time-lapse imaging and morphological data, leading to improved implantation success rates. These models integrate multiple factors such as patient age, hormone levels, ovarian response, and embryo quality, allowing for highly personalized pregnancy predictions and enabling clinicians to optimize treatment plans and medication dosages. In 2019, Blank et al. utilized random forest models to improve blastocyst implantation predictions, achieving enhanced accuracy by incorporating patient data [17]. Griesinger et al. (2020) applied non-invasive deep learning methods to predict embryo ploidy using time-lapse imaging, further advancing embryo selection strategies [18]. Similarly, Drakeley et al. (2021) developed the ERICA algorithm, a deep learning tool that ranks embryos based on their genetic health (ploidy status) and likelihood of implantation, providing a non-invasive alternative to traditional biopsy methods [19]. More recent studies have demonstrated the power of deep learning in predicting live birth outcomes. Huang et al. (2022) analyzed over 10,000 embryo samples, achieving an impressive AUC of 0.968 for live birth predictions using time-lapse data, thereby automating the process of embryo evaluation and ranking [20]. Sawada et al. (2021) similarly developed an AI system that used the Attention Branch Network to analyze over 140,000 embryo images, successfully predicting

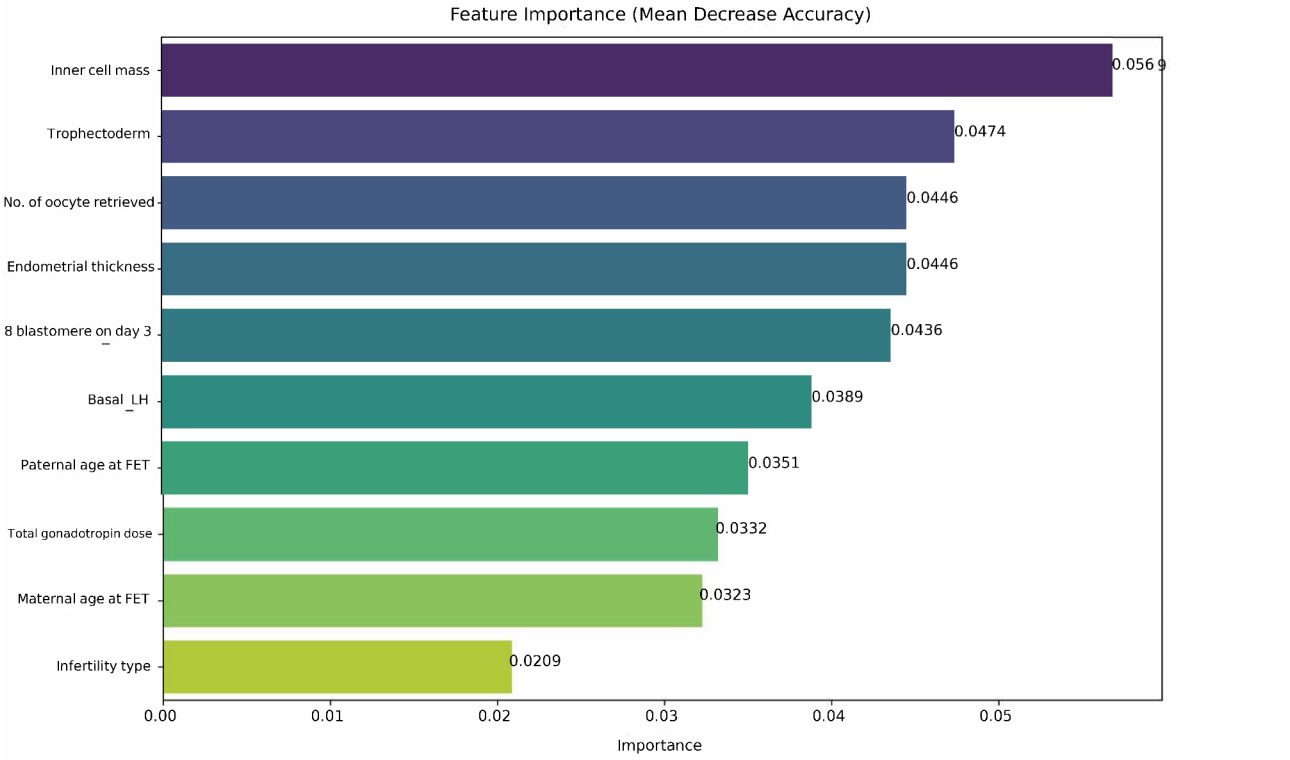

Fig. 3 Ranking of feature importance using Mean Decrease Accuracy with the extra trees model

live birth probabilities based on a confidence score threshold [21]. Additionally, Hassan et al. (2020) introduced a machine learning model that combined feature selection with several ML techniques, including random forests and support vector machines, identifying key IVF-related factors such as maternal age and embryo transfer day. This model demonstrated a significantly higher accuracy in predicting IVF pregnancy success compared to traditional methods [12]. These advancements underscore the transformative role of ML in ART, offering clinicians more precise tools for improving IVF success rates, although further validation in clinical settings is needed to ensure reliability and explainability.

Principal findings

His study successfully identified key factors influencing live birth rate following single vitrified-warmed blastocyst transfer (SVBT) in women of advanced maternal age (AMA). By employing 19 machine learning algorithms and comprehensive data processing methods, significant predictive factors were determined, including inner cell mass quality, trophectoderm quality, number of oocytes retrieved, endometrial thickness, and blastocyst derived from 8 blastomeres on day 3. The top three models, Stacking Classifier, Extra Trees Classifier, and Random Forest Classifier, demonstrated superior predictive performance compared to traditional statistical methods, achieving higher Area Under the Curve (AUC) values and balanced sensitivity and specificity.

Results in the context of what is known

The identified significant predictive factors align with existing literature. Inner cell mass quality and trophectoderm quality have been consistently highlighted as critical components for embryo implantation and development [22]. High-quality inner cell mass embryos are more likely to implant successfully and develop into healthy fetuses, underscoring the importance of inner cell mass quality in IVF treatments [9, 23]. Similarly, high-quality trophectoderm is essential for successful embryo implantation and early development. It facilitates attachment to the endometrium and improves pregnancy rates, emphasizing the significance of embryo health before transfer [23, 24]. Additionally, blastocysts derived from 8 blastomeres on day 3 are considered well-developed and have high implantation potential. This finding is consistent with previous studies on embryo developmental potential, indicating the embryo's health and developmental potential during the early stages, which are vital for predicting pregnancy outcomes [25]. The number of

oocytes retrieved is known to reflect ovarian response. A higher number of oocytes increases the chances of selecting high-quality embryos, thereby enhancing the likelihood of pregnancy [26]. This variable is significant as it affects both the quantity and quality of embryos available for transfer [27]. Optimal endometrial thickness is necessary for embryo implantation, as both excessively thin and thick endometria can lead to implantation failure or early miscarriage. An optimal thickness (usually between 8 and 14 mm) provides a conducive environment for the embryo to implant and develop, making this an important predictive factor [9, 28].

Identifying these significant predictive factors not only enhances our understanding of the critical elements for SVBT success but also provides valuable insights for clinicians to optimize treatment protocols and improve live birth rate.

Clinical implications

The findings of this study provide valuable insights for clinicians in optimizing IVF treatment protocols for advanced maternal age patients. High-quality inner cell mass and trophectoderm are crucial for improving the success rate of SVBT, indicating that greater attention should be paid to embryo quality assessment in clinical practice. Additionally, when selecting blastocysts for transfer, the condition of the blastocyst during the cleavage stage should be fully considered. Priority should be given to blastocysts derived from 8 blastomeres due to their better developmental potential. A higher number of oocytes retrieved and appropriate endometrial thickness are also important factors for success. By focusing on these identified predictive factors, clinicians can improve live birth rate following SVBT cycles.

Research implications

The superior performance of machine learning models over traditional methods suggests that integrating advanced analytical techniques into clinical practice can enhance predictive accuracy and support more personalized patient care. However, these findings need to be validated in larger, multi-center studies to ensure their confident application in clinical settings.Future research should incorporate more variables and utilize larger, multi-center datasets to further improve the predictive accuracy and generalizability of the models. Further studies should also explore the application of machine learning models in other aspects of IVF treatments to enhance overall treatment efficacy and patient outcomes. Finally, research should focus on the integration of real-time data analysis and predictive modeling into clinical workflows to provide timely and actionable insights for patient management.

Strengths and limitations

The strengths of this study include the use of multiple machine learning algorithms and rigorous data processing methods, leading to robust and unbiased predictive models. Notably, this study focuses on SVBT outcomes in advanced maternal age (AMA) patients, addressing a gap in the literature and providing valuable insights for optimizing clinical treatment protocols. Our models demonstrated high predictive accuracy, with the Stacking Classifier, Extra Trees Classifier, and Random Forest Classifier achieving AUCs of 0.791, 0.784, and 0.768, respectively, highlighting their strong performance and feasibility for clinical application.

However, several limitations must be acknowledged. First, to address class imbalance, we employed feature selection and ensemble learning methods to minimize the risks of overfitting and misclassification often associated with oversampling techniques such as SMOTE. It is important to note that synthetic samples may not always accurately represent the minority class [29]. Second, while the models showed high predictive accuracy, their performance needs further validation in larger, multi-center datasets to ensure broader applicability. Third, the analysis focused solely on vitrified-thawed blastocyst transfers, excluding fresh transfers, which are crucial when considering cumulative live birth rates (cLBR). The omission of fresh transfers, a significant aspect of ART outcomes, should be recognized. Fourth, some patients contributed multiple cycles, which may introduce bias as certain individuals were included more than once. Although the study was cycle-focused rather than patient-centered, this could impact data independence. Lastly, the study did not account for all possible factors affecting SVBT outcomes, such as genetic and environmental influences, which may limit the comprehensiveness of the models.

Future research should enhance the generalizability of the findings by incorporating larger, multi-center datasets that represent diverse populations and regions. Additionally, analyzing both vitrified-thawed and fresh blastocyst transfers will offer a more comprehensive understanding of ART outcomes, including cumulative live birth rates. Furthermore, future models should consider a wider range of factors, including genetic predispositions, lifestyle influences, and environmental factors, to improve predictive power and model comprehensiveness.

Conclusions

This study identified key predictive factors for live birth rates following single vitrified-warmed blastocyst transfer (SVBT) in women of advanced maternal age (AMA), including inner cell mass quality, trophectoderm quality, number of oocytes retrieved, endometrial thickness, and blastocysts from 8-cell stage embryos on day 3. The Stacking, Extra Trees, and Random Forest classifiers

demonstrated the highest predictive accuracy, outperforming traditional models. These findings can aid in managing patient expectations and enhance clinical decision-making. Future studies should validate these results across multi-center cohorts and explore additional factors, such as fresh embryo transfers.

Author contributions
Lidan Liu and Yihua Yang designed the study. Lidan Liu and Bo Liu conducted the experiments and data collection. Ming Liao and Qiuying Gan performed the data analysis. Lidan Liu wrote the main manuscript text. Qianyi Huang prepared Figs. 1, 2 and 3. All authors reviewed and approved the final manuscript.

Funding
This research was supported by a self-funded project from the Health and Family Planning Commission of Guangxi Zhuang Autonomous Region, grant number Z-A20230585.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Ethical approval
The study protocol obtained ethical approval from the Institutional Review Boards of the participating hospitals, specifically the First Affiliated Hospital of Guangxi Medical University Medical Ethics Committee (Approval Number: 2024-E445-01) and the Nanning Maternity and Child Health Hospital Medical Ethics Committee (Approval Number: YX20240627-1). Informed consent was secured from all participants.The authors confirm that the guidelines approved by the local institution were followed.
This study was a retrospective dual-center analysis of data from two reproductive medicine centers.

Author details
[1]Guangxi Reproductive Medical Center, The First Affiliated Hospital of Guangxi Medical University, Nanning, China
[2]Reproductive Center, Nanning Maternity and Child Health Hospital, Nanning, Guangxi, China

Received: 27 July 2024 / Accepted: 2 October 2024
Published online: 07 October 2024

References

1. WHO W. 1 in 6 people globally affected by infertility. World Health Organization. *Retrieved from WHO Website.* 2023.
2. Saket Z, Källén K, Lundin K, Magnusson Å, Bergh C. Cumulative live birth rate after IVF: trend over time and the impact of blastocyst culture and vitrification. Hum Reprod Open. 2021;2021(3):hoab021. https://doi.org/10.1093/hropen/hoab021.
3. Önalan G, Tunç M, Tohma A, Günakan E, Eryılmaz T, Zeyneloğlu HB. Extending the culture of cleavage-stage embryos to the blastocyst stage after warming increases the chance of live birth: does it have a regenerative effect? Arch Gynecol Obstet. 2023;307(6):1969–74. https://doi.org/10.1007/s00404-023-07031-7.
4. Ferraretti AP, Nygren K, Andersen AN, et al. Trends over 15 years in ART in Europe: an analysis of 6 million cycles†. Hum Reprod Open. 2017;2017(2):hox012. https://doi.org/10.1093/hropen/hox012.
5. Sokol P, Clua E, Pons MC, et al. Developing and validating a prediction model of live birth following single vitrified-warmed blastocyst transfer. Reprod Biomed Online. 2024;49(1):103890. https://doi.org/10.1016/j.rbmo.2024.103890.
6. Ubaldi FM, Cimadomo D, Vaiarelli A, et al. Advanced maternal age in IVF: still a challenge? The Present and the future of its treatment. Front Endocrinol. 2019;10:94. https://doi.org/10.3389/fendo.2019.00094.
7. Chen Y, Zhou J, Chen Y, et al. Pregnancy outcomes after Frozen Embryo Transfer and fresh embryo transfer in women of advanced maternal age: single-center experience. J Clin Med. 2022;11(21):6395. https://doi.org/10.3390/jcm11216395.
8. Temmesen CG, Faber Frandsen T, Svarre-Nielsen H, Petersen KB, Clemensen J, Andersen HLM. Women's reflections on timing of motherhood: a meta-synthesis of qualitative evidence. Reprod Health. 2023;20(1):30. https://doi.org/10.1186/s12978-022-01548-x.
9. Liu X, Chen Z, Ji Y. Construction of the machine learning-based live birth prediction models for the first in vitro fertilization pregnant women. BMC Pregnancy Childbirth. 2023;23(1):476. https://doi.org/10.1186/s12884-023-05775-3.
10. Barnett-Itzhaki Z, Elbaz M, Butterman R, et al. Machine learning vs. classic statistics for the prediction of IVF outcomes. J Assist Reprod Genet. 2020;37(10):2405–12. https://doi.org/10.1007/s10815-020-01908-1.
11. Cai J, Jiang X, Liu L, et al. Pretreatment prediction for IVF outcomes: generalized applicable model or centre-specific model? Hum Reprod. 2024;39(2):364–73. https://doi.org/10.1093/humrep/dead242.
12. Hassan MR, Al-Insaif S, Hossain MI, Kamruzzaman J. A machine learning approach for prediction of pregnancy outcome following IVF treatment. Neural Comput Appl. 2020;32(7):2283–97. https://doi.org/10.1007/s00521-018-3693-9.
13. Ogutu JO, Piepho HP, Schulz-Streeck T. A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* 2011;5(S3):S11. https://doi.org/10.1186/1753-6561-5-S3-S11
14. De Mello MEF, Da Silva Rocha É, De Morais FL et al. Exploring the Performance of Machine Learning Models and Predictive Factors for Fetal Death: Preliminary Results. In: Onwubiko C, Rosati P, Rege A, eds. *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media*. Springer Proceedings in Complexity. Springer Nature Singapore; 2024:3–24. https://doi.org/10.1007/978-981-99-6974-6_1
15. Peluso C, Oliveira RD, Laporta GZ, et al. Are ovarian reserve tests reliable in predicting ovarian response? Results from a prospective, cross-sectional, single-center analysis. Gynecol Endocrinol. 2021;37(4):358–66. https://doi.org/10.1080/09513590.2020.1786509.
16. Gardner DK, Schoolcraft WB. Culture and transfer of human blastocysts. Curr Opin Obstet Gynaecol. 1999;11(3):307–11. https://doi.org/10.1097/00001703-199906000-00013.
17. Blank C, Wildeboer RR, DeCroo I, et al. Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. Fertil Steril. 2019;111(2):318–26. https://doi.org/10.1016/j.fertnstert.2018.10.030.
18. Griesinger G, Cometti B, Trevisan S, Reply. Endometrial thickness performs poorly as a predictor of IVF outcome, but does the endometrial receptivity array perform any better? Hum Reprod Open. 2020;2020(2). https://doi.org/10.1093/hropen/hoaa015.
19. Drakeley A, Flores-Saiffe A, Chavez-Badiola A, et al. P–244 ERICA's (embryo ranking Intelligent classification Assistant) ranking, based on ploidy prediction, is strongly correlated with pregnancy outcomes. Hum Reprod. 2021;36(Supplement1). https://doi.org/10.1093/humrep/deab130.243.
20. Huang B, Zheng S, Ma B, Yang Y, Zhang S, Jin L. Using deep learning to predict the outcome of live birth from more than 10,000 embryo data. BMC Pregnancy Childbirth. 2022;22(1):36. https://doi.org/10.1186/s12884-021-04373-5.
21. Sawada Y, Sato T, Nagaya M, et al. Evaluation of artificial intelligence using time-lapse images of IVF embryos to predict live birth. Reprod Biomed Online. 2021;43(5):843–52. https://doi.org/10.1016/j.rbmo.2021.05.002.
22. Shear MA, Vaughan DA, Modest AM, et al. Blasts from the past: is morphology useful in PGT-A tested and untested frozen embryo transfers? Reprod Biomed Online. 2020;41(6):981–9. https://doi.org/10.1016/j.rbmo.2020.07.014.
23. Qiu J, Li P, Dong M, Xin X, Tan J. Personalized prediction of live birth prior to the first in vitro fertilization treatment: a machine learning method. J Transl Med. 2019;17(1):317. https://doi.org/10.1186/s12967-019-2062-5.
24. Wang CW, Kuo CY, Chen CH, Hsieh YH, Su ECY. Predicting clinical pregnancy using clinical features and machine learning algorithms in in vitro fertilization. Viganò P. ed PLOS ONE. 2022;17(6):e0267554. https://doi.org/10.1371/journal.pone.0267554.

25. Wu J, Zhang J, Kuang Y, Chen Q, Wang Y. The effect of Day 3 cell number on pregnancy outcomes in vitrified-thawed single blastocyst transfer cycles. Hum Reprod. 2020;35(11):2478–87. https://doi.org/10.1093/humrep/deaa209.
26. Vermey BG, Chua SJ, Zafarmand MH, et al. Is there an association between oocyte number and embryo quality? A systematic review and meta-analysis. Reprod Biomed Online. 2019;39(5):751–63. https://doi.org/10.1016/j.rbmo.2019.06.013.
27. Gorelova IV, Galkina JA, Kuznetsova LV, Rulev MV, Zazerskaya IE. The number of oocytes as a predictor of pregnancy rates in the cycles of assisted reproductive technology. Transl Med. 2020;7(1):53–8. https://doi.org/10.18705/2311-4495-2020-7-1-53-58.
28. Li L, Cui X, Yang J, Wu X, Zhao G. Using feature optimization and Light-GBM algorithm to predict the clinical pregnancy outcomes after in vitro fertilization. Front Endocrinol. 2023;14:1305473. https://doi.org/10.3389/fendo.2023.1305473.
29. Alkhawaldeh IM, Albalkhi I, Naswhan AJ. Challenges and limitations of synthetic minority oversampling techniques in machine learning. World J Methodol. 2023;13(5):373–8. https://doi.org/10.5662/wjm.v13.i5.373.

Publisher's note