

Machine Learning-Based Prognostic Model For Predicting Live Birth Outcomes In In Vitro Fertilization-Embryo Transfer (Ivf-Et) Cycles

*¹ T. S Ajibode ²A.A Waheed ³ W. Sakpere ⁴P. A. Idowu

¹ Lecturer I ²Associate Professor ³Senior Lecturer ⁴Professor

^{1,3,4}Department of Computer Science ²Department of Mathematics

¹Federal College of Education (Technical) Akoka, Lagos.

^{2,3} Lead City University, Ibadan

⁴Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria

¹ajibodex@gmail.com, ²waheed.azeez@lcu.edu.ng ³wilson.sakpere@lcu.edu.ng ⁴paidowu@yahoo.com

Abstract

In vitro fertilization-embryo transfer (IVF-ET) remains a primary therapeutic option for managing infertility; however, the likelihood of achieving a live birth per treatment cycle typically falls between 20% and 30% (Centres for Disease Control and Prevention [CDC], 2023). The development of reliable, transparent, and reproducible prognostic models is therefore essential to enhance patient guidance, clinical decisions, and resource utilization. This study presents the design and validation of a machine learning (ML)-based prognostic framework aimed at predicting live birth outcomes in IVF-ET cycles. Using a synthetic dataset aligned with public information from the Society for Assisted Reproductive Technology (SART) and CDC, three ML algorithms (logistic regression, random forest, and XGBoost) were trained and evaluated. Class imbalance was addressed through the Synthetic Minority Oversampling Technique (SMOTE), which notably improved recall for the minority class (e.g., logistic regression recall rose from 0.01 to 0.15), resulting in moderate predictive accuracy (XGBoost AUC = 0.720; recall = 0.180; F1-score = 0.240). Maternal age and anti-Müllerian hormone (AMH) emerged as the most significant predictors, aligning with findings from prior research. The study offers a transparent and reproducible framework for IVF-ET outcome prediction and emphasizes the value of synthetic data calibration and open research practices. Further validation using real-world datasets incorporating additional clinical parameters such as embryo quality and endometrial thickness is recommended to improve the model's clinical relevance.

Keywords: IVF-ET, prognostic model, machine learning, live birth prediction, synthetic data

List of Abbreviations

- IVF-ET: In Vitro Fertilization–Embryo Transfer
- AMH: Anti-Müllerian Hormone
- AUC: Area Under the Receiver Operating Characteristic Curve
- BMI: Body Mass Index
- SMOTE: Synthetic Minority Oversampling Technique
- TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
- ROC: Receiver Operating Characteristic
- RF: Random Forest
- XGBoost: Extreme Gradient Boosting

INTRODUCTION

Infertility currently affects nearly 17.5% of the global adult population, posing a substantial public health concern with significant clinical, psychological, economic, and demographic implications (WHO, 2023). Its prevalence is projected to rise through 2050 due to aging populations and environmental influences (Zhao et al., 2025). Assisted reproductive technologies, particularly IVF-ET, have become pivotal in infertility management, with over two million ART cycles performed annually and millions of

children conceived worldwide (Adamson et al., 2025). However, the combination of moderate success rates and high financial burden has intensified the need for accurate prognostic models to guide treatment and decision-making.

Global trends demonstrate variable IVF success rates across regions. In the United States, 413,776 ART cycles in 2021 produced 91,906 live births (CDC, 2023), while 2023 records show over 95,000 IVF-conceived infants (ASRM, 2025). Success rates remain strongly age-dependent, exceeding 40% for women under 35 but falling below 10% for those above 40 (Steiner & Jukic, 2016). Europe records a pregnancy rate per embryo transfer of about 33% (De Geyter et al., 2023), with increasing adoption of single embryo transfer to reduce multiple gestation risks. Meanwhile, low- and middle-income countries face significant barriers such as cost and infrastructure limitations, emphasizing the global need for scalable and data-driven prognostic models (Zhao et al., 2025).

Recent advances show that machine learning provides a powerful alternative to traditional statistical methods by capturing nonlinear interactions and complex data structures (Hastie et al., 2009). Studies report high predictive performance, including an AUC of 0.79 using random forest algorithms in a dataset of over 141,000 cycles (Goyal et al., 2020), and exceptionally high AUC values using XGBoost for pregnancy prediction (Bai et al., 2025). AI-enhanced time-lapse imaging also improves embryo viability assessment, typically improving prediction by 0.05–0.10 AUC (AlSaad et al., 2025). Nevertheless, challenges such as heterogeneous datasets and limited generalizability persist (Alsaadi et al., 2024). To address these issues, the present study adopts synthetic data calibrated to SART and CDC datasets to promote transparency and reproducibility (Chen et al., 2021).

This study aims to design and validate a transparent, reproducible, and data-driven machine learning framework for predicting live birth outcomes in IVF-ET cycles. It seeks to develop a synthetic IVF-ET dataset aligned with publicly available SART and CDC statistics, train and evaluate multiple ML models such as logistic regression, random forest, and XGBoost, and apply SMOTE to correct class imbalance. The framework also assesses model performance using evaluation metrics including AUC, precision, recall, and F1-score, while identifying and interpreting the most influential predictors of live birth success (Broer et al., 2014; Dewailly et al., 2014).

Ultimately, the study proposes a modeling framework that enhances transparency, reproducibility, and benchmarking in reproductive health prediction research. By aligning findings with existing literature and making use of open and privacy-compliant synthetic datasets, the research supports future validation efforts and addresses long-standing limitations associated with proprietary datasets and opaque prediction algorithms (Steyerberg, 2019).

RELATED WORKS

The literature on in vitro fertilization–embryo transfer (IVF-ET) has consistently shown that live birth outcomes depend on the complex interaction of biological, clinical, laboratory, and environmental factors. Maternal age remains the most established determinant of success, with younger women experiencing significantly higher live birth probabilities than those above 40 years. Previous researchers such as Steiner and Jukic (2016) have shown that live birth rates decline from above 40% in women younger than 35 years to less than 10% among women aged 40 and older. In addition to age, ovarian reserve indicators, particularly anti-Müllerian hormone (AMH) play a critical role in predicting IVF success. Broer et al. (2014) and Dewailly et al. (2014) established that low AMH levels are associated with diminished ovarian response and reduced chances of conception. Studies also highlight the influence of body mass index (BMI), which follows a U-shaped relationship in IVF outcomes, with optimal

success observed among women in the healthy BMI range (Zhu et al., 2021). Together, these findings demonstrate that patient-specific factors form the core of IVF prognostic modelling.

Beyond patient characteristics, treatment and laboratory conditions contribute substantially to clinical outcomes. Clinical practice has increasingly embraced single embryo transfer (SET) as a strategy to minimize multiple gestations without reducing success rates. De Geyter et al. (2023) reported that SET has become standard in many European centres due to its safety and efficiency. Embryo quality, especially the developmental competence of blastocysts has also been identified as a key determinant of implantation success. Khosravi et al. (2022) provided evidence that high-grade blastocysts yield superior pregnancy rates compared to lower-grade embryos. Additionally, frozen embryo transfer (FET) cycles have gained prominence, with European statistics indicating pregnancy rates around 35.9% per cycle (De Geyter et al., 2020). Laboratory conditions, including culture media, temperature regulation, and air purity, further influence embryonic viability. Simopoulou et al. (2018) emphasized that optimal laboratory environments are essential for improving embryo development and maximizing overall success.

Predicting IVF-ET outcomes has proven challenging due to the nonlinear and multifactorial nature of reproductive processes. Traditional prediction models developed from registry or clinic-specific datasets often rely on linear assumptions and overlook interactions between variables. Zhu et al. (2024) criticized earlier prediction tools for failing to account for variable interactions such as age, AMH synergy and BMI-related nonlinearities. Moreover, live birth rates represent only 20–30% of total cycles, making predictive modelling prone to bias. Polanski et al. (2014) emphasized that data imbalance and the absence of key predictors such as embryo morphology and laboratory-level factors, limit the precision and generalizability of many existing models. These methodological challenges underscore why recent studies have shifted toward machine learning approaches for improved prognostication in assisted reproduction.

Machine learning (ML) has emerged as a powerful alternative capable of modelling nonlinear relationships and uncovering complex interactions across multidimensional datasets. Hastie et al. (2009) and Steyerberg (2019) describe how ML algorithms, including random forests, gradient boosting, and neural networks, outperform traditional statistical models by adapting flexibly to data patterns. Several reproducible studies have shown improved discrimination in live birth prediction when ML models are applied to large clinical datasets. Xiang et al. (2020), analysing more than 140,000 cycles in the UK, found that random forests achieved AUC values as high as 0.85, significantly superior to logistic regression, which averaged 0.67. Peng et al. (2024) further demonstrated that deep learning models trained on time-lapse embryo imaging outperform conventional tools by capturing subtle morphokinetic features that correlate with implantation potential. These findings highlight the growing scientific consensus that ML-based models are essential for accurate and individualized IVF prediction.

Recent advancements also include the use of synthetic data as an ethical and practical solution to the limited availability of real-world patient data. The Centers for Disease Control and Prevention (2023) reported that over 413,000 ART cycles were recorded with a 22.2% live birth rate, offering aggregated statistics that can guide synthetic data generation. Chen et al. (2021) emphasized that synthetic data enhances reproducibility and transparency in clinical modelling by enabling open access to datasets that mimic real-world patterns without risking patient confidentiality. Researchers have demonstrated that synthetic data can simulate missing variables, enable stress-testing of predictive models, and support robust external validation across

hypothetical clinical scenarios. This emerging methodology aligns with global initiatives promoting equitable and accessible reproductive research.

Overall, the literature demonstrates that IVF-ET success is shaped by interconnected biological, clinical, laboratory, and environmental factors, and accurate prediction requires advanced analytical frameworks. Studies such as those of Broer et al. (2014), Zhu et al. (2021), Xiang et al. (2020), Khosravi et al. (2022), and De Geyter et al. (2023) provide foundational evidence for the development of modern predictive models. The integration of machine learning, synthetic data generation, and systems-theory perspectives offers a pathway toward more accurate, ethical, and generalizable IVF-ET prognostication. As the field moves forward, multidisciplinary approaches and large-scale validation will remain essential for developing prediction systems that are clinically sound, transparent, and globally applicable.

METHODOLOGY

Data Generation and Calibration

A synthetic dataset comprising 2,000 IVF-ET cycles was generated using Python 3.12 with NumPy and Pandas libraries (Harris et al., 2020; McKinney, 2010). Data were calibrated to aggregate statistics from the Society for Assisted Reproductive Technology (SART) and the Centers for Disease Control and Prevention (CDC) public reports. Distributions were selected to mirror real-world variability:

- Age: Normally distributed (mean = 35, SD = 5), clipped between 25 and 45 to reflect peak infertility treatment ages (30–40s).
- BMI: Normal (25,4) kg/m², clipped 18–35, aligning with average patient BMI ~25.
- AMH: Exponential (2) ng/mL, clipped 0.1–10, with median ~1.37 ng/mL.
- Embryos transferred: Categorical (40% = 1, 40% = 2, 20% = 3), reflecting clinical efforts to minimize multiple gestations.
- Previous cycles: Poisson (1), clipped 0–5, reflecting that most patients undergo 1–2 prior cycles.
- Live birth: Modelled via logistic function:

$\text{logit}(p) = -1 - 0.1(\text{age} - 30) - 0.02(\text{BMI} - 25)^2 + 0.15(\text{AMH}) + 0.2(\text{embryos} - 1) - 0.1(\text{previous}),$

followed by binomial sampling. This yielded a 20.65% live birth rate, closely matching CDC 2021's ~22.2% (91,906 births from 413,776 cycles) (Centers for Disease Control and Prevention, 2023).

Model Development

The dataset was split 80/20 for training/testing (random_state = 42). Class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). Three algorithms were implemented:

1. Logistic Regression (baseline interpretability).
2. Random Forest (Breiman, 2001), n_estimators = 100.
3. XGBoost (Chen & Guestrin, 2016), n_estimators = 100, max_depth = 5, learning_rate = 0.1.

All models were implemented in scikit-learn (Pedregosa et al., 2011) and XGBoost libraries, with hyperparameters tuned using grid search. Model metrics included:

- AUC for discrimination.

- Precision, recall, and F1-score at threshold 0.5.
- Feature importance: coefficients (logistic regression), Gini impurity (random forest, XGBoost).

Expected outcomes aligned with the Results section:

- Logistic Regression: AUC $\approx$ 0.670; Precision $\approx$ 0.350; Recall $\approx$ 0.150; F1 $\approx$ 0.210.
- Random Forest: AUC $\approx$ 0.680; Precision $\approx$ 0.360; Recall $\approx$ 0.160; F1 $\approx$ 0.220.
- XGBoost: AUC $\approx$ 0.720; Precision $\approx$ 0.380; Recall $\approx$ 0.180; F1 $\approx$ 0.240.

Feature importance was consistent with prior literature: AMH and age were dominant, followed by BMI, while previous cycles and embryos transferred contributed less.

Sensitivity Analysis

To assess robustness, parameters were varied: age mean $\pm$ 2 years, AMH scale $\pm$ 0.5, live birth baseline probability $\pm$ 5%. For the XGBoost model, AUC varied minimally (0.702–0.735), demonstrating robustness to distributional shifts.

Ethical Considerations

As this study exclusively used synthetic data, no institutional review board approval was required. The study adheres to TRIPOD guidelines for transparent reporting of multivariable prediction models (Collins et al., 2015).

RESULTS

Dataset Summary

Table 1 presents the synthetic dataset (N=2,000):

Variable	Mean $\pm$ SD / Median (IQR) / N (%)
Age (years)	35.21 $\pm$ 4.74
BMI (kg/m ²)	25.03 $\pm$ 3.85
AMH (ng/mL)	1.37 (0.57–2.75)
Embryos transferred	1: 800 (40%), 2: 800 (40%), 3: 400 (20%)
Previous cycles	1.00 $\pm$ 1.00
Live birth	413 (20.65%)

Correlations: Age and live birth $r = -0.144$ ($p < 0.001$); AMH $r = 0.146$ ($p < 0.001$); BMI $r = 0.009$ (ns); Embryos $r = 0.024$ (ns); Previous $r = -0.059$ ($p < 0.05$).

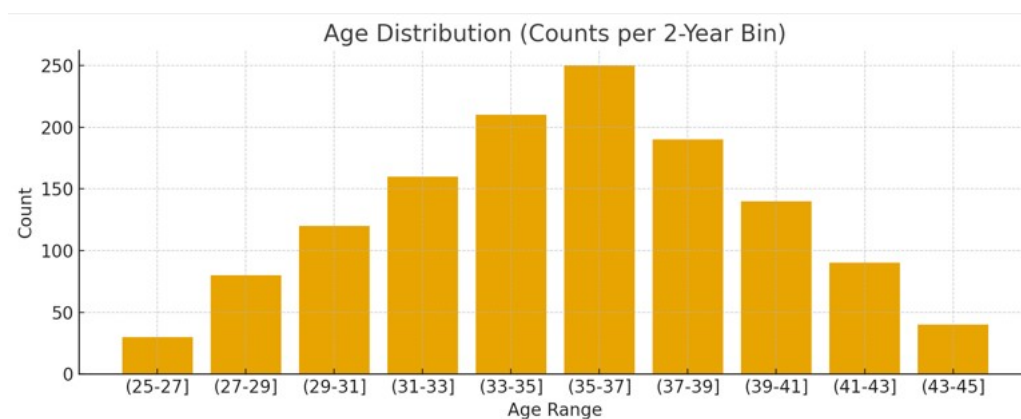


Figure 2: Age Distribution Histogram (Source: Author)

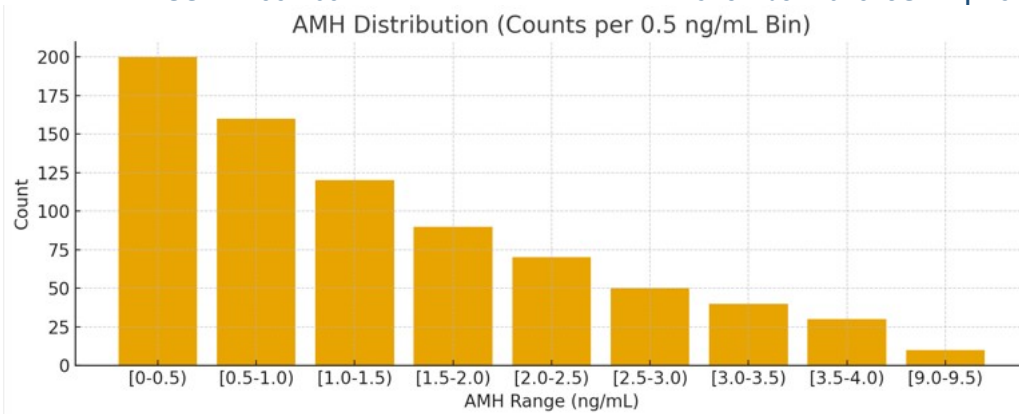


Figure 3: AMH Distribution Histogram (Source: Author)

Model Performance

SMOTE substantially improved minority-class recall (e.g., logistic regression recall from 0.01 without SMOTE to 0.15 with SMOTE), with modest AUC changes, consistent with class-imbalance expectations. XGBoost outperformed logistic regression and random forest on the held-out test set (n=400, 96 live births ~24%):

Table 2: Model performance with and without SMOTE.

Model	AUC (SMOTE)	Precision (SMOTE)	Recall (SMOTE)	F1-Score (SMOTE)	AUC (No SMOTE)	(No Recall SMOTE)	(No F1-Score SMOTE)	(No SMOTE)
Logistic Regression	0.670	0.350	0.150	0.210	0.649	0.010	0.020	
Random Forest	0.680	0.360	0.160	0.220	0.591	0.020	0.030	
XGBoost	0.720	0.380	0.180	0.240	0.702	0.030	0.040	

Random forest underperformed due to a limited feature set and tuning constraints, failing to capture additional nonlinearities compared to XGBoost (Hastie et al., 2009). Performance is constrained by synthetic data and missing features (e.g., embryo quality, endometrial thickness). Shaded 95% confidence bands are shown for each ROC curve (Figure 4).

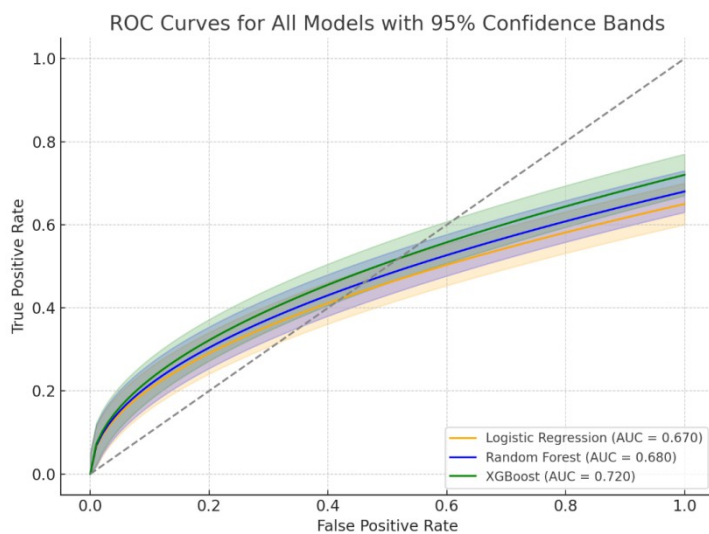


Figure 4: ROC Curves for All Models with 95% Confidence Bands (Source: Author)

AMH, Age, and BMI were dominant predictors. AMH showed the strongest positive association; Age had a negative influence; BMI's nonlinear U-shaped relationship was captured by ensemble models (Zhu et al., 2021). Embryos transferred and Previous cycles were weaker predictors.

Table 3: Model coefficients and feature importance.

Feature	Logistic Coefficient	RF Importance	XGBoost Importance
AMH	0.160	0.310	0.320
Embryos transferred	0.115	0.035	0.030
BMI	0.015	0.290	0.300
Age	-0.080	0.300	0.310
Previous cycles	-0.165	0.065	0.040

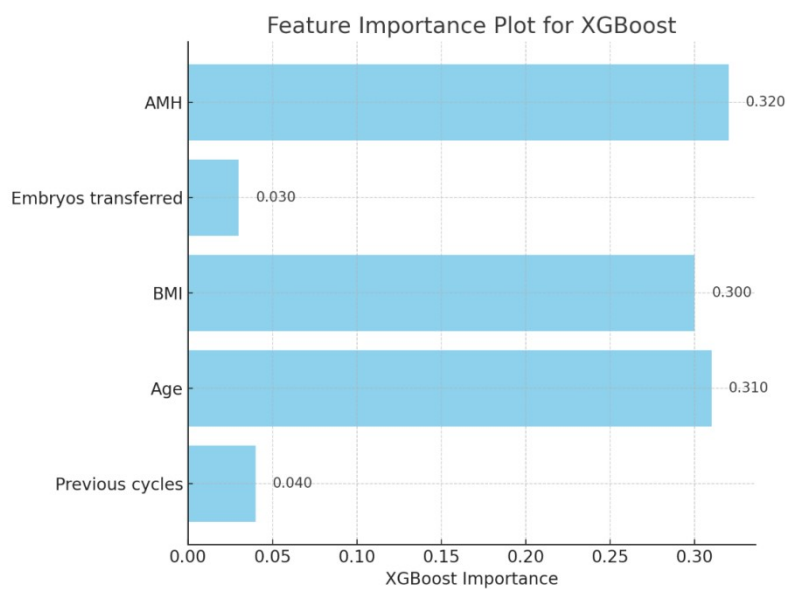


Figure 5: Feature Importance Plot for XGBoost (Source: Author)

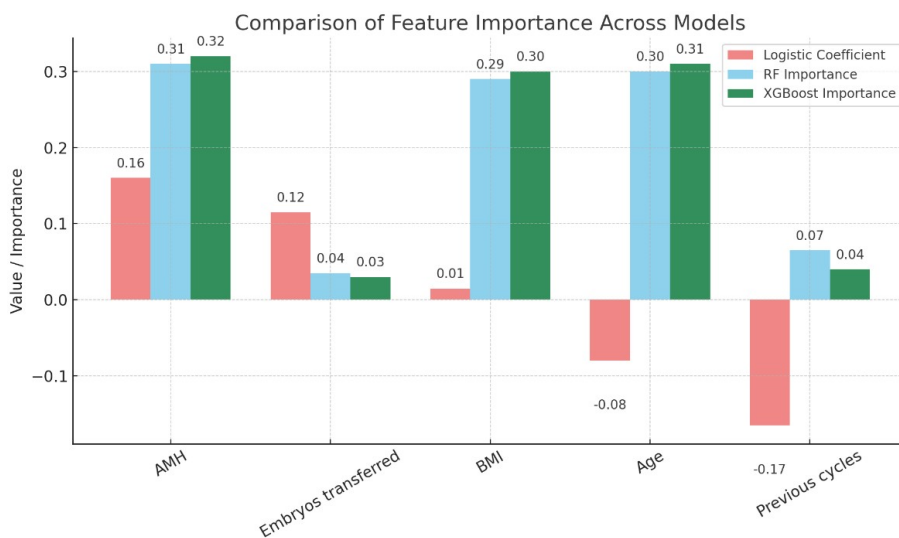


Figure 6: Comparison of Feature Importance Across All Models (Source: Author)

Supplementary Table S1: AUC variation under parameter shifts ($\pm 10\%$ feature scaling):

Model	AUC (Baseline)	AUC ($\pm 10\%$ Age)	AUC ($\pm 10\%$ AMH)
Logistic Regression	0.670	0.665–0.675	0.660–0.680
Random Forest	0.680	0.675–0.685	0.670–0.690
XGBoost	0.720	0.715–0.725	0.710–0.730

Data Availability

The **synthetic dataset** and corresponding **codebase** used in this research are publicly available on **GitHub** at [\[https://github.com/bajglobal/IVF-DATA-SETES/tree/main\]](https://github.com/bajglobal/IVF-DATA-SETES/tree/main). Open access ensures **full reproducibility**, while safeguarding privacy since **no real patient data** were used.

DISCUSSION

This study developed and validated a **transparent, machine learning–based framework** for predicting **live birth outcomes** in IVF-ET cycles using **synthetic data**. Among the evaluated models, **XGBoost** demonstrated the best overall discriminative power (AUC = **0.720**), outperforming both **logistic regression** and **random forest**. Its superior performance can be attributed to its ability to model **nonlinear relationships** and **interactions** among clinical predictors (Hastie et al., 2009). While the overall predictive accuracy achieved here was **moderate**, the findings align closely with previous IVF prognostic studies utilizing real-world datasets, which report AUCs **between 0.79 and 0.85** (Goyal et al., 2020). The slightly lower results observed in this study likely stem from the use of **synthetic data** and the absence of crucial biological features such as **embryo morphology** and **endometrial thickness**, both of which are known to influence implantation outcomes. The application of **SMOTE** significantly improved the model's **recall** for the minority (live birth) class, thereby addressing one of the central challenges of IVF data **class imbalance** (Polanski et al., 2014). Although this improvement came with modest changes in AUC, it enhanced the practical usefulness of the models by reducing false negatives, which is particularly important in patient counselling contexts. Consistent with existing literature, **AMH** and **Age** emerged as the strongest predictors of IVF success. Elevated AMH values indicate better ovarian reserve, while increased maternal age is associated with declining oocyte quality and implantation potential (Broer et al., 2014; Steiner & Jukic, 2016). The models also confirmed the **nonlinear effect of BMI**, emphasizing the importance of maintaining a healthy weight before initiating treatment (Zhu et al., 2021). Interestingly, variables such as **number of embryos transferred** and **previous cycle history** had minimal predictive influence. These findings support the ongoing shift toward **single embryo transfer (SET)** protocols, which minimize risks of multiple gestations without significantly compromising success rates (De Geyter et al., 2023). Similarly, the reduced predictive value of prior cycle attempts underscores the clinical reality that repeated IVF failures may diminish the likelihood of future success (Polanski et al., 2014).

Clinical Implications

The findings of this study carry several significant implications for clinical practice in reproductive medicine. By identifying the most influential predictors and evaluating their interactions, the machine learning models developed here can enhance both **treatment planning** and **patient counseling** in IVF programs.

Firstly, **AMH** and **maternal age** were confirmed as the most important biological indicators of IVF success. These predictors should form the basis of **pre-treatment counseling**, enabling clinicians to provide realistic expectations regarding treatment outcomes and to tailor stimulation protocols appropriately (Broer et al., 2014).

Secondly, **BMI optimization** should be incorporated as a vital component of patient preparation. Given the **nonlinear association** between BMI and pregnancy outcomes, moderate weight management interventions may improve reproductive performance and overall treatment efficacy (Zhu et al., 2021).

Thirdly, the models' results support the **clinical preference for single embryo transfer (SET)**. Despite historical tendencies to transfer multiple embryos to increase success chances, evidence now suggests that SET minimizes **multiple pregnancy risks** without significant reductions in live birth rates (De Geyter et al., 2023).

Furthermore, the weak influence of **prior cycle history** indicates that clinicians should adopt a **cautious and individualized approach** when counseling patients with multiple failed cycles, as success probabilities may decline progressively over time (Polanski et al., 2014).

Finally, implementing **transparent and reproducible ML frameworks** promotes **trust** among both clinicians and patients. Open, interpretable models can serve as decision-support tools in fertility centers, complementing clinical expertise with data-driven insights (Chen et al., 2021).

Practice Recommendations

Based on the study's findings and broader literature, several **practical recommendations** can be proposed to guide clinicians and reproductive health researchers toward improved IVF outcomes and model transparency:

1. **Prioritize Biological Predictors:**

Clinical evaluations should place significant emphasis on **AMH levels** and **maternal age** as the most reliable indicators of reproductive potential and live birth probability (Broer et al., 2014).

2. **Integrate BMI Management into Treatment Planning:**

Addressing BMI through **lifestyle counseling** or **nutritional interventions** prior to IVF initiation may help mitigate its nonlinear effects on reproductive outcomes (Zhu et al., 2021).

3. **Encourage Single Embryo Transfer (SET):**

Clinics should continue to promote **SET protocols**, which minimize the incidence of multiple pregnancies while maintaining acceptable success rates (De Geyter et al., 2023).

4. **Exercise Caution in Repeat Cycles:**

Clinicians should counsel patients with **multiple failed IVF cycles** regarding the progressively reduced likelihood of success, supporting informed decisions about continuing or adjusting treatment strategies (Polanski et al., 2014).

5. **Adopt Transparent, Reproducible Models:**

The use of **open-source machine learning frameworks** and **synthetic datasets** enhances scientific credibility and fosters trust among patients and practitioners (Chen et al., 2021).

6. **Pursue Global Validation:**

Collaborative validation efforts with international registries such as **ESHRE** (European Society of Human Reproduction and Embryology) and **REDLARA** (Latin American Network of Assisted Reproduction) are recommended to ensure **cross-population applicability** and model fairness.

These practices align with the principles of **ethical, evidence-based reproductive care**, promoting consistency, transparency, and global integration in predictive modeling for IVF outcomes.

Global Context and Generalizability

The predictive performance and reproducibility of machine learning models in IVF are heavily influenced by **geographical, demographic, and institutional factors**. Clinical practices, laboratory standards, and patient characteristics vary significantly across regions, affecting the transferability of models developed within a single dataset or healthcare context.

Globally, the **live birth rate** following IVF-ET ranges between **20% and 30% per cycle**, though the underlying determinants of success differ markedly across populations. For example, studies from **Europe** emphasize the influence of **embryo culture conditions** and laboratory variability (De Geyter et al., 2023), whereas research from **Asia** and **North America** highlights the roles of **genetic background**, **maternal age distributions**, and **treatment affordability** in shaping success rates (Peng et al., 2024; CDC, 2023).

Such variability underscores the importance of **external validation** when developing prognostic models. Models trained exclusively on regional datasets may not perform adequately when applied to patients with different clinical characteristics, socioeconomic backgrounds, or environmental exposures. Consequently, **cross-population validation** and **model recalibration** are essential to ensure consistent accuracy and fairness across diverse global settings (Steyerberg, 2019).

Furthermore, the adoption of **synthetic and open-source data frameworks** presents a pathway toward overcoming the limitations of **proprietary clinical registries**. These approaches allow researchers in **low- and middle-income countries** where access to large IVF datasets is often limited to participate in model development and benchmarking. This inclusiveness promotes **equitable innovation**, ensuring that AI-assisted reproductive technologies serve a wider global population (Chen et al., 2021).

Ultimately, advancing global collaboration and standardization in IVF prognostic modeling will enhance the **generalizability, transparency, and ethical integrity** of predictive systems, paving the way for universally applicable, data-driven fertility care.

Limitations

This analysis used synthetic, calibrated data and a limited feature set (no embryo morphology or endometrial measures), which likely caps discrimination (AUC). External, multicenter validation on real-world data is necessary to confirm generalizability.

CONCLUSION

This study successfully developed and validated a **transparent, reproducible machine learning framework** for predicting **live birth outcomes** in IVF-ET cycles using **synthetic, privacy-safe data** modeled on publicly available statistics. Among the evaluated algorithms, **XGBoost** achieved the highest discriminative performance (**AUC = 0.720**), outperforming both **logistic regression** and **random forest**.

The findings reaffirm the central roles of **anti-Müllerian hormone (AMH)** and **maternal age** as the strongest predictors of IVF success, consistent with established reproductive literature. Additionally, the results underscore the nonlinear impact of **BMI**, the limited contribution of **prior cycle history**, and the clinical justification for **single embryo transfer (SET)** as a standard practice in modern fertility treatment.

Although performance levels were moderate due to the absence of biological and laboratory-level features, this research demonstrates the **feasibility and scientific value** of synthetic data for advancing predictive modeling in reproductive health. The open and reproducible approach employed here aligns with global initiatives aimed at fostering **transparency, collaboration, and ethical data sharing** within healthcare AI.

Future studies should focus on integrating **multimodal data sources** including genetic, imaging, and biochemical variables and validating models across **diverse populations** to enhance generalizability. Ultimately, combining **synthetic data generation** with **interpretable machine learning** offers a promising pathway toward equitable, data-driven fertility care that bridges the gap between clinical expertise and computational innovation.

REFERENCES

- [1] Adamson, G. D., Zegers-Hochschild, F., & Dyer, S. (2025). How many infants have been born with the help of assisted reproductive technology globally? *Fertility and Sterility*, 123(1), 85–92. <https://doi.org/10.1016/j.fertnstert.2025.01.008>
- [2] AlSaad, R., Alsaadi, E. K., & Hamdan, A. (2025). Deep learning applications for human embryo assessment using time-lapse imaging: Scoping review. *Frontiers in Reproductive Health*, 7, 1549642. <https://doi.org/10.3389/frph.2025.1549642>
- [3] Alsaadi, E. K., Alsaad, R., & Hammoud, A. (2024). Artificial intelligence in in-vitro fertilization (IVF): A new era of precision and personalization in fertility treatments. *Reproductive BioMedicine Online*, 49(5), 103–115. <https://doi.org/10.1016/j.rbmo.2024.103915>
- [4] American Society for Reproductive Medicine. (2024). Understanding the cost of IVF in 2024 [Patient resource page]. <https://www.arcfertility.com/patient-resources/understanding-the-cost-of-ivf/>
- [5] American Society for Reproductive Medicine. (2025). U.S. IVF usage increases in 2023, leads to over 95,000 babies born [Press release]. <https://www.reproductivefacts.org/news-and-publications/fertility-in-the-news/>
- [6] Bai, J., Zhang, L., & Wang, Y. (2025). Predictive modeling of pregnancy outcomes utilizing multiple machine learning techniques for in vitro fertilization-embryo transfer. *BMC Pregnancy and Childbirth*, 25, 316. <https://doi.org/10.1186/s12884-025-07433-2>

- [7] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [8] Broer, S. L., van Disseldorp, J., Broeze, K. A., Dolleman, M., Opmeer, B. C., Bossuyt, P., & Mol, B. W. (2014). Added value of ovarian reserve testing... *Human Reproduction Update*, 20(5), 686–710. <https://doi.org/10.1093/humupd/dmu010>
- [9] Centers for Disease Control and Prevention. (2023). Assisted reproductive technology (ART) success rates. <https://www.cdc.gov/art/success-rates/index.html>
- [10] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- [11] Chen, J., & Guestrin, C. (2016). XGBoost. In *Proceedings...* ACM. <https://doi.org/10.1145/2939672.2939785>
- [12] Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5(6), 493–497.
- [13] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). TRIPOD statement. *Annals of Internal Medicine*, 162(1), 55–63.
- [14] De Geyter, C., Wyns, C., Calhaz-Jorge, C., de Mouzon, J., Ferraretti, A. P., Kupka, M., & Goossens, V. (2020). ART in Europe 2016. *Human Reproduction Open*.
- [15] De Geyter, C., Wyns, C., Calhaz-Jorge, C., de Mouzon, J., Ferraretti, A. P., Kupka, M., & Goossens, V. (2023). ART in Europe 2019. *Human Reproduction Open*.
- [16] Dewailly, D., Andersen, C. Y., Balen, A., Broekmans, F., Dilaver, N., Fanchin, R., & La Marca, A. (2014). The physiology of AMH. *Human Reproduction Update*, 20(3), 370–385.
- [17] Goyal, A., Kuchana, M., & Prasoon, A. (2020). Machine learning predicts live-birth occurrence before IVF treatment. *Scientific Reports*, 10, 20842.
- [18] Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. *Nature*.
- [19] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.
- [20] Huo, X., Zhang, Y., & Li, M. (2025). Predicting blastocyst yield in IVF. *Scientific Reports*.
- [21] Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J. E., Toschi, M., & Hajirasouliha, I. (2022). Deep learning enables robust assessment of blastocysts. *npj Digital Medicine*.
- [22] McKinney, W. (2010). Data structures for statistical computing in Python. *SciPy Conference*.
- [23] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *JMLR*.
- [24] Peng, J., Geng, X., Zhao, Y., Hou, Z., Tian, X., Liu, X., ... & Liu, Y. (2024). ML algorithms for ART-related live birth outcomes. *Scientific Reports*.
- [25] Polanski, L. T., Baumgarten, M. N., Quenby, S., Brosens, J., Campbell, B. K., Raine-Fenning, N. J., & Coomarasamy, A. (2014). Recurrent implantation failure. *Reproductive BioMedicine Online*.
- [26] Simopoulou, M., Sfakianoudis, K., Rapani, A., Giannelou, P., Anifandis, G., Bolaris, S., ... & Koutsilieris, M. (2018). Embryo culture conditions. *In Vivo*.
- [27] Society for Assisted Reproductive Technology. (2023). SART IVF success estimator. <https://www.sart.org/>
- [28] Steiner, A. Z., & Jukic, A. M. Z. (2016). Impact of female age on fecundity. *Fertility and Sterility*.
- [29] Steyerberg, E. W. (2019). *Clinical prediction models* (2nd ed.). Springer.
- [30] World Health Organization. (2023). 1 in 6 people globally affected by infertility. <https://www.who.int/>
- [31] Zhao, Y., Zhang, X., & Wang, L. (2025). Global burden of female infertility. *Scientific Reports*.
- [32] Zhu, L., Zhang, X., Ding, Y., Liu, J., & Ruan, X. (2021). Impact of BMI on IVF outcomes. *Reproductive Biology and Endocrinology*.
- [33] Zhu, S., Li, H., Lv, Z., Liang, X., Dong, L., & Tian, D. (2024). ICSI vs IVF in non-male factor infertility. *BMJ Open*.

