Journal of Public Health Science (JoPHS)
Volume 3, No 2 – Juni 2026
e-ISSN :3047-3632

MACHINE LEARNING MODELS FOR PREDICTING LIVE BIRTH IN WOMEN WITH PCOS UNDERGOING IVF/ICSI: A SYSTEMATIC REVIEW

Cora Angela[1], Sarma Nursani Lumbanraja[2]
[1]Faculty of Medicine, University of North Sumatera, Medan, Indonesia
[2]Department of Obstetrics and Gynaecology, Medical Faculty, University of North Sumatera, Medan, Indonesia
coraangela88@gmail.com

Received: 28-05- 2026	Revised: 10-06-2026	Approved: 22-06-2026

ABSTRACT

Introduction. Polycystic Ovary Syndrome (PCOS) merupakan salah satu gangguan endokrin paling umum pada wanita usia reproduktif, dengan prevalensi sekitar 6–20% tergantung kriteria diagnostik yang digunakan. PCOS ditandai oleh disfungsi ovulasi, hiperandrogenisme, dan gangguan metabolik seperti resistensi insulin yang berkontribusi terhadap infertilitas. Banyak pasien PCOS memerlukan Assisted Reproductive Technology (ART), termasuk IVF dan ICSI, namun outcome utama yaitu live birth masih sulit diprediksi karena dipengaruhi banyak faktor. Perkembangan Artificial Intelligence (AI) dan Machine Learning (ML) menawarkan pendekatan baru dalam prediksi outcome klinis, namun sebagian besar model masih memiliki keterbatasan metodologis dan kurangnya validasi eksternal. ***Method.*** *Penelitian ini merupakan systematic review yang bertujuan mengidentifikasi dan mengevaluasi model machine learning untuk memprediksi live birth pada pasien PCOS yang menjalani IVF/ICSI. Pencarian literatur dilakukan pada PubMed, ScienceDirect, dan OpenAlex sesuai pedoman PRISMA 2020.* ***Result.*** *Dari 115 studi yang diidentifikasi, tiga studi memenuhi kriteria inklusi. Model machine learning yang dilaporkan meliputi XGBoost, Random Forest, dan model berbasis regresi logistik. XGBoost menunjukkan performa terbaik dengan AUC 0,822, Random Forest dengan AUC 0,60, dan model regresi logistik sekitar AUC 0,80. Faktor prediktor yang konsisten meliputi usia, BMI, durasi infertilitas, kualitas embrio, dan parameter hormonal. Namun, seluruh studi masih terbatas pada validasi internal tanpa validasi eksternal yang memadai.* ***Conclusion.*** *Model machine learning menunjukkan potensi dalam memprediksi live birth pada pasien PCOS yang menjalani IVF/ICSI, namun belum terdapat model yang benar-benar superior dan tervalidasi secara luas. Diperlukan pengembangan model dengan validasi eksternal multi-center serta evaluasi risiko bias yang lebih ketat sebelum implementasi klinis.*

Keywords*: Machine Learning Models; Predicting Live Birth; Women With Pcos; Undergoing Ivf/Icsi*

INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is one of the most common endocrine disorders in women of reproductive age, with a prevalence of about 6–20% depending on the diagnostic criteria used.[4] PCOS is characterized by ovulation dysfunction, hyperandrogenism, and metabolic disorders such as insulin resistance, which contribute to infertility in this population.[5]

Therefore, many women with PCOS require Assisted Reproductive Technology (ART) interventions, including In Vitro Fertilization (IVF) and Intracytoplasmic Sperm Injection (ICSI), to achieve pregnancy. Although IVF/ICSI has increased the chances of pregnancy in PCOS patients, the most important clinical outcome of live birth rates is still difficult to accurately predict because it is influenced by many complex factors such as age, body mass index, ovarian reserve, embryo quality, and response to ovarian stimulation. The non-linear interaction between these variables makes conventional

statistical approaches such as logistic regression have limitations in capturing the diversity of reproductive data.

In recent years, the development of Artificial Intelligence (AI) and Machine Learning (ML) has provided a new approach in predicting complex clinical outcomes. Algorithms such as Random Forest, Support Vector Machine, Gradient Boosting, and Neural Networks are able to capture non-linear patterns and multidimensional interactions in IVF clinical data that traditional models cannot handle.[6 7]

Several studies have shown that machine learning models can be used to predict IVF outcomes with promising performance, including clinical pregnancy and live birth. However, most of these models still have methodological limitations. In addition, many models were developed using data from a single clinical center, thus limiting the generalization of outcomes to a wider population, particularly PCOS patients who have heterogeneous metabolic and hormonal characteristics.

Furthermore, most of the current literature still focuses on the prediction of clinical pregnancy, while live birth as the most important clinical outcome in assisted reproduction is still relatively underexplored in the context of machine learning. To date, there has been no consensus encapsulating the performance and methodological quality of AI models specifically developed to predict live birth in women with PCOS undergoing IVF/ICSI. Therefore, systematic evidence is needed to identify, evaluate, and summarize the machine learning models used in predicting live births in PCOS patients undergoing IVF/ICSI.

METHODOLOGY

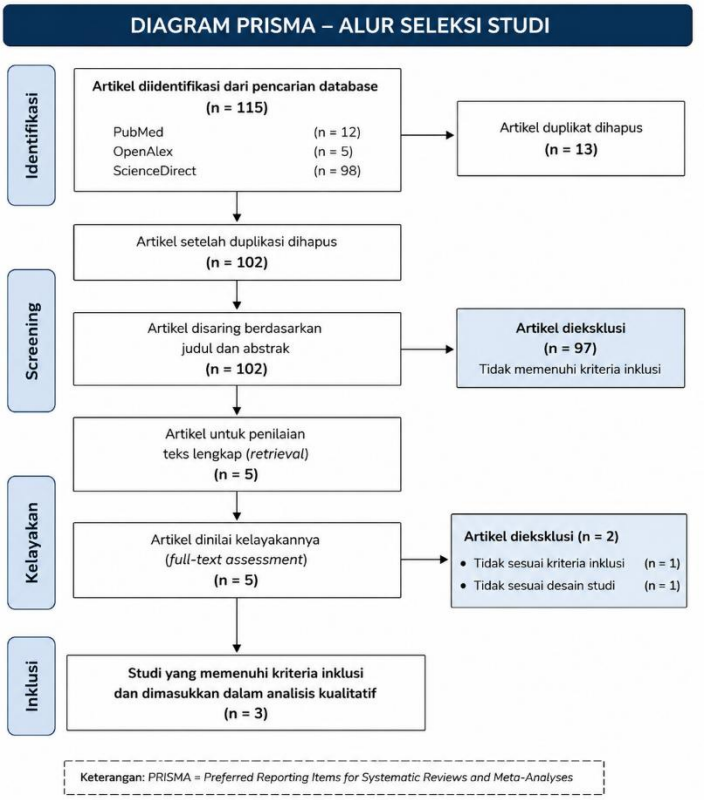
This study is a systematic review that aims to identify and evaluate a machine learning-based prediction model to predict live birth outcomes in women with Polycystic Ovary Syndrome (PCOS) who undergo in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI). This study was prepared following the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 to ensure transparency and completeness of reporting.

Literature searches were conducted on several electronic databases, namely PubMed, ScienceDirect, and OpenAlex, from the beginning of the database to the date of the last search. The search strategy uses a combination of keywords related to PCOS, IVF/ICSI, machine learning/artificial intelligence, and live birth. All relevant studies are then selected in stages through the elimination of duplication, title and abstract screening, and full-text assessment based on the predetermined inclusion and exclusion criteria. The selection process was carried out by two researchers independently, and differences were resolved through discussion.

The included studies include retrospective and prospective cohort research, registry-based studies, and machine learning-based prediction model development and validation studies. The population studied was women with a diagnosis of PCOS based

on international criteria such as Rotterdam, NIH, or AE-PCOS who underwent IVF or ICSI. Studies that only used traditional statistical models without a machine learning component, as well as publications in the form of reviews, editorials, or case reports, were not included in the study.

The primary outcome in this review was live birth, while secondary outcomes included clinical pregnancy, ongoing pregnancy, miscarriage, and implantation rate. Data were extracted using a standard form based on the CHARMS framework, including study characteristics, population, type of machine learning algorithm, predictor variables, validation methods, and model performance such as AUC, sensitivity, specificity, and model calibration.



RESULTS

Of the 115 studies obtained, three studies were found that met the research inclusion criteria. A total of three relevant retrospective cohort studies were identified and analyzed in this review, all of which focused on the development and validation of machine learning models to predict reproductive outcomes in PCOS patients undergoing IVF/ICSI. The study by Suqin Zhu, et al (2025) in China is a retrospective cohort study for the development and validation of seven machine learning models conducted on 1,062 PCOS patients diagnosed under the Rotterdam Criteria or China's guidelines for the diagnosis and treatment of PCOS, with a median age of 31 years, a majority BMI of <28 kg/m^2, and 55.08% having experienced infertility ≥3 years.

Interventions include ART (IVF and ICSI), fresh cycle, with ovarian stimulation protocol using an antagonist protocol (GnRH antagonist), and embryonic stage including cleavage-stage embryo or blastocyst. The outcomes assessed were live birth, namely pregnancy that reached the age of ≥28 weeks with vital signs after childbirth and clinical pregnancy based on transvaginal ultrasound. Of the 1,062 embryo transfer cycles, 466 cycles (43.88%) resulted in live births. The machine learning models tested included Decision Tree, K-nearest neighbors, LightGBM, Naive Bayes, Random Forest, SVM, and XGBoost. The results showed that XGBoost was the best model with an AUC of 0.822 (95% CI: 0.777–0.867), accuracy of 0.752, sensitivity of 0.772, specificity of 0.732, and Brier Score of 0.172. The SHAP analysis identified seven major factors that affect live birth, including the number of embryos transferred, the type of embryo transfer, the mother's age, the duration of infertility, BMI, testosterone levels, and progesterone levels on the day of triggering. Decision Curve Analysis shows that the XGBoost model provides a higher net benefit in the threshold probability range of 10%–90%.

Research by Payam Amini, et al (2021) in Iran used a cohort design to compare six machine learning methods in predicting successful births in 6,071 IVF/ICSI cycles, with 5,105 cycles involving PCOS patients and 1,199 cycles in patients without PCOS. The outcome assessed was a successful delivery confirmed through the presence of an intrauterine gestational sac. The models tested included Random Forest, Support Vector Machine, XGBoost, Logistic Regression, Naïve Bayes, and Linear Discriminant Analysis. The results showed Random Forest as the best model with an AUC of 0.60 (95% CI: 0.55–0.64), accuracy of 0.81, sensitivity of 0.82, specificity of 0.39, PPV of 0.99, and NPV of 0.01. Logistic Regression has the lowest performance with an accuracy of 0.64. Other models such as SVM, XGBoost, Naïve Bayes, and LDA perform similarly in sensitivity and PPV that are almost as high as RF, but RF still excels in total accuracy. Influential predictive factors include embryo count, oocyte count, causes of infertility, age, and PCOS. Patients with PCOS showed a higher rate of successful delivery than non-PCOS (82.1% vs 78.8%).

The study by Suqin Zhu et al (2024) in China used a retrospective cohort design with the aim of developing and validating a predictive nomogram model for live birth in 1,242 PCOS patients based on the Rotterdam criteria or Chinese clinical guidelines. Interventions involved IVF (73.51%) and ICSI (26.49%) with fresh embryo transfer and GnRH antagonist protocols. The majority of embryos were in the cleavage stage (88.24%) and the rest were blastocysts (11.76%). The outcome assessed was live birth which was defined as a live birth of ≥28 weeks with vital signs. The model was developed using multivariable logistic regression with Boruta feature selection (Random Forest-based feature selection method). The model showed an AUC of 0.804 at training and 0.807 at validation, with good calibration (Hosmer-Lemeshow p=0.103 and p=0.102). Negative risk factors included age, duration of infertility, BMI ≥28 kg/m^2, high testosterone, and

high progesterone on hCG days, while positive factors included blastocyst transfer, number of embryos transferred, and embryo quality.

DISCUSSION

The findings of the three studies show that the machine learning model has promising potential in predicting live birth in PCOS patients undergoing IVF/ICSI. The XGBoost and Random Forest algorithms show better predictive performance than classic methods such as logistic regression. This shows that the machine learning approach is able to capture the non-linear relationships and complex interactions between clinical variables that play a role in the success of IVF.

A study by Suqin Zhu et al (2025) shows that XGBoost has the best performance with an AUC of 0.822, and shows good calibration ability and clinical net benefit through decision curve analysis. In addition, SHAP analysis provides a clear interpretation of the clinical factors that most affect live birth, such as the number of embryos transferred, maternal age, BMI, and hormonal factors. These findings suggest that machine learning models not only provide accurate predictions but can also provide relevant clinical interpretability.

Meanwhile, a study by Payam Amini et al (2021) showed that although Random Forest was the best-performing model, the relatively low AUC value (0.60) showed limitations in the model's discriminating ability on a mixed PCOS and non-PCOS dataset. This indicates that population heterogeneity can significantly affect model performance, as well as emphasizing the importance of developing more specific models in the PCOS subpopulation. The research of Suqin Zhu et al (2024) strengthens the finding that a logistic regression-based model strengthened by the Boruta technique can still achieve good performance with an AUC of around 0.80, especially when good internal validation is carried out.

However, limitations in the absence of external validation suggest that model generalization is still a major challenge in clinical implementation. Overall, all three studies showed consistency that factors such as maternal age, BMI, duration of infertility, embryo quality, and reproductive hormone levels were important predictors of live birth. However, the differences in performance between models suggest that there is no truly superior and widely validated machine learning model for use in routine clinical practice. In addition, most models are still limited to internal validation and have not been tested externally in different populations, thus limiting the generalization of results.

CONCLUSION

Based on three retrospective cohort studies analyzed, the machine learning model showed good potential in predicting live births in women with PCOS undergoing IVF/ICSI, specifically XGBoost and Random Forest. Clinical factors such as maternal age,

BMI, duration of infertility, quality and number of embryos transferred, and hormonal profiles (testosterone and progesterone) consistently play an important role in determining the chances of a successful live birth. However, all reported models still have major methodological limitations, mainly due to the lack of external validation in independent populations, so the model's generalization ability to different clinical settings remains uncertain. In addition, most studies still rely on internal validation and have not fully met the reporting and evaluation standards of comprehensive prediction models. Therefore, the development of more robust machine learning models with multi-center external validation, better calibration reporting, and rigorous bias risk evaluation is needed before the model can be widely implemented in assisted reproductive clinical practice in PCOS patients.

DAFTAR PUSTAKA

1. Zhu S, Huang Z, Chen X, Jiang W, Zhou Y, Zheng B, et al. Construction and evaluation of machine learning-based prediction model for live birth following fresh embryo transfer in IVF/ICSI patients with polycystic ovary syndrome. J Ovarian Res. 2025;18:70. doi:10.1186/s13048-025-01654-x.
2. Amini P, Ramezanali F, Parchehbaf-Kashani M, Maroufizadeh S, Omani-Samani R, Ghaheri A. Factors associated with in vitro fertilization live birth outcome: A comparison of different classification methods. Int J Fertil Steril. 2021;15(2):128-134. doi:10.22074/IJFS.2020.134582 .
3. Zhu S, Chen X, Li R, Jiang W, Zheng B, Sun Y. Constructing a predictive model for live birth following fresh embryo transfer in antagonist protocol for polycystic ovary syndrome. J Assist Reprod Genet. 2024;41(10):2709-2719. doi:10.1007/s10815-024-03232-4.
4. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod. 2004;19(1):41-47. doi:10.1093/humrep/deh098.
5. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al.; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Fertil Steril. 2018;110(3):364-379. doi:10.1016/j.fertnstert.2018.05.004.
6. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29. doi:10.1038/s41591-018-0316-z.
7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.