

# Predicting cumulative live birth rate for patients undergoing *in vitro* fertilization (IVF)/intracytoplasmic sperm injection (ICSI) for tubal and male infertility: a machine learning approach using XGBoost

Zhiyan Chen, Duoduo Zhang, Jingran Zhen, Zhengyi Sun, Qi Yu

Department of Obstetrics and Gynecology, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China.

*To the Editor:* Confronted with increasing infertility worldwide, this decade has witnessed a sharp rise in the utilization of artificial reproduction technology (ART).<sup>[1]</sup> Nevertheless, the success rate of *in vitro* fertilization (IVF)/intracytoplasmic sperm injection (ICSI) depends on many factors. Moreover, patients receiving ART not only incur high expenses but also an increased risk of severe side effects, such as ovarian hyperstimulation syndrome, infection, and multiple pregnancies.<sup>[2]</sup> Consequently, the accurate prediction of ART outcomes has attracted tremendous interest.

A considerable number of logistic regression models have been developed to predict ovarian stimulation,<sup>[3]</sup> pregnancy outcomes,<sup>[4]</sup> and adverse obstetric outcomes.<sup>[5]</sup> Although prediction models based on traditional statistical methods have been broadly applied, their clinical utility is hindered by their low predictive efficacy. Therefore, more accurate models to predict IVF/ICSI outcomes are needed.

The accelerating development of computer technology heightened the popularization of artificial intelligence in medicine. These new machine learning methods are distinguished with their enhanced performance in comparison to the conventional methods. As one of the well-received machine learning techniques, eXtreme Gradient Boosting (XGBoost), has been gradually put into medical use and recognized for its remarkable capacity to mine data.<sup>[6-8]</sup> XGBoost, a decision-tree-based algorithm, proved to be the best algorithm for machine learning in a prediction competition hosted by Kaggle.com.<sup>[9]</sup>

To better predict pregnancy outcomes and provide patient-tailored counseling and management for IVF/ICSI, we designed this study and selected XGBoost to build the prediction model. The aim of this study was twofold. First, we sought to develop a new prediction model based on the

machine learning method to predict IVF/ICSI outcomes. Second, by comparing the performance of the machine learning model and the conventional logistic regression model, we attempt to determine the efficacy of the machine learning model.

We retrospectively studied continuous patients' data from the Peking Union Medical College Hospital, China, from July 2014 to March 2018. The patients were referred to the hospital for IVF/ICSI treatment with tubal or male infertility. All data were de-identified and relabeled with unique patients' identifier codes. Exclusion criteria included donor oocyte or sperm use; patients with endometriosis or endocrine diseases, such as hyperandrogenism, diabetes, or thyroid diseases; and patients with missing data. The study was approved by the Institutional Review Board of Peking Union Medical College Hospital (S-K829).

The clinical characteristics, sex hormone levels, and controlled ovarian hyperstimulation (COH) features of the IVF/ICSI cycles were used as model construction variables. Each patient's age, body mass index (BMI), infertility type, infertility duration, and COH protocol (gonadotropin-releasing hormone antagonist [GnRH-a] long protocol, GnRH-a ultra-long protocol, GnRH-a short protocol, GnRH antagonist protocol, and mini-stimulation protocol) were extracted from the medical records. The serum levels of the sex hormones (human follicle-stimulating hormone [FSH], estrogen [E2], luteinizing hormone [LH], prolactin [PRL], and testosterone [T]) were collected at two time points during the IVF/ICSI cycle (basal: 0; the second day after trigger: 1). Live birth, defined as giving birth to live newborns at >28 weeks of gestation, was the primary study outcome. We evaluated the cumulative outcome of each patient, including the first

## Access this article online

Quick Response Code:



Website:  
www.cmj.org

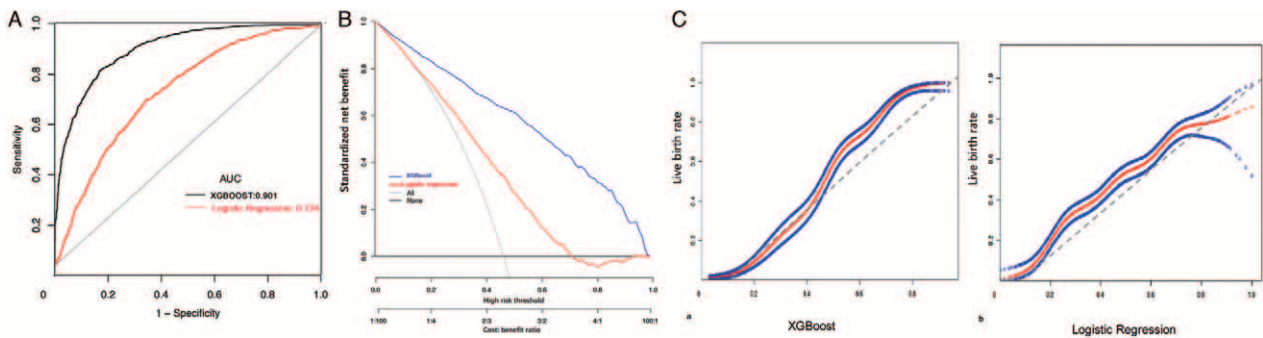
DOI:  
10.1097/CM9.0000000000001874

**Correspondence to:** Prof. Qi Yu, Prof. Zhengyi Sun, Department of Obstetrics and Gynecology, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China  
E-Mail: yuqi2008001@sina.com, sunzhengyi@263.net

Copyright © 2022 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2022;135(8)

Received: 21-09-2021; Online: 08-12-2021 Edited by: Yanjie Yin



**Figure 1:** (A) The ROC curves of the two models. The XGBoost model presented a higher discriminatory power than the conventional one (AUC [conventional]: 0.724, 95% CI 0.708–0.741; AUC [XGBoost]: 0.901, 95% CI 0.890–0.912;  $P < 0.001$ ). (B) The DCA of the two models. The DCA curve of the XGBoost model was above that of the conventional model with a greater range on the axes, indicating that the net benefit of the XGBoost model was larger than the conventional model. (C) The calibration curves of the two models. Both models presented a good calibration of the probability of live birth. AUC: Area under the curve; CI: Confidence interval; DCA: Decision curve analysis; ROC: Receiver operating curves; XGBoost: eXtreme Gradient Boosting.

fresh cycle and all subsequent freeze-thaw cycles from the same ovarian stimulation and identified them as live and no live births.

Statistical analyses were performed using R (<http://www.R-project.org>) and EmpowerStats software 2.2 (<http://www.empowerstats.com>, X&Y Solutions, Inc., Boston, MA). We built a conventional logistic regression model based on multivariate logistic regression analysis. A backward stepwise variable selection procedure with bootstrap resampling was applied to select the variables. And then the open-source XGBoost package was applied to analyze feature importance and acquired the probability threshold of live births. The predictive efficacy of the two models was evaluated by measuring their sensitivity, specificity, positive predictive value, and negative predictive value. Receiver operating curves and corresponding area under the curve (AUC) values of the two models were compared. Then, the calibration curves were assessed. The decision curve analysis (DCA) was performed to compare the clinical utility of the models.

A total of 3012 patients were included in the model construction with 2101 IVF and 911 ICSI cases. The patients' clinical characteristics, sex hormone levels, and COH features are listed as two groups (live birth and no live birth) in Supplementary Table 1, <http://links.lww.com/CM9/A835>.

The top vital features selected by the XGBoost model were age, estrogen levels on the second day after trigger (E21), PRL levels on the second day after trigger (PRL1), basal LH levels (LH0), LH levels on the second day after trigger (LH1), E20 (basal estrogen levels), basal PRL levels (PRL0), total consumption of FSH. And the contribution of each feature to the model construction is illustrated in Supplementary Table 2, <http://links.lww.com/CM9/A835> and Supplementary Figure 2, <http://links.lww.com/CM9/A835>.

The features of the conventional model, selected by backward stepwise analysis, included age, secondary infertility, ICSI, No. of previous IVF, total consumption of FSH, FSH0, T0, PRL1, LH1, E21, P1, and T1 [Supplementary Table 3, <http://links.lww.com/CM9/A835>].

The predictive performance of the two models is presented in Supplementary Table 4, <http://links.lww.com/CM9/A835> and Figure 1. Compared with the conventional logistic regression model, the XGBoost model had a higher AUC value, which represented a better discriminatory power (AUC [conventional]: 0.724, 95% confidence interval [CI] 0.708–0.741; AUC [XGBoost]: 0.901, 95% CI 0.890–0.912;  $P < 0.001$ ). Good calibration was observed for the probability of live birth in both models. The DCA curve of the XGBoost model was greater than that of the conventional model, indicating a larger net benefit of the XGBoost model.

In this study, to predict the IVF/ICSI outcomes for patients with tubal or male infertility, we built an XGBoost model that showed higher performance and better discriminative capacity than a conventional logistic regression model. According to the DCA results, the XGBoost model also had a larger net benefit than the conventional model, indicating its better clinical potential.

For patients receiving ART, an accurate prediction of the success rate and subsequent individualized treatment strategies can be beneficial.<sup>[10]</sup> Numerous studies have been conducted to estimate the chance of live birth. The most popular model is the McLernon model, based on the UK national data of 184,269 complete cycles since 1991, providing a personalized estimated cumulative chance of live births before treatment and after the first fresh embryo transfer (C-index of 0.72–0.73). However, some imperative factors that might be potential predictors for live birth, such as anti-Müllerian hormone and BMI, were not included.<sup>[11,12]</sup> In addition, compared with the previous studies using traditional statistical methods, the XGBoost model in our study showed better performance with a high AUC value of 0.901.

Among various machine learning algorithms, XGBoost is one of the most clinically recognized because of its remarkable predictive ability. XGBoost is a decision-tree-based algorithm that collects multiple decision trees to improve its classification capabilities.<sup>[13]</sup> The decision trees are developed by selecting the most discriminative features from the feature candidate pool, allowing the

classifier to interact directly with the features. Hence, XGBoost can assemble models with a weak predictive ability or missing data and demonstrate excellent prediction competence to solve complicated problems.<sup>[14]</sup> Similar to our study, others have constructed prediction models with XGBoost and have confirmed its higher clinical value compared to the traditional methods.<sup>[15]</sup> Qiu *et al*<sup>[16]</sup> used XGBoost algorithm for personalized prediction of live births for IVF/ICSI patients, with an AUC of 0.73. Amini *et al*<sup>[17]</sup> used different machine learning approaches to predict the probability of successful delivery and suggested that random forests had the best performance (AUC = 0.81). Compared with our study, none of the aforementioned studies verified the superiority of the machine learning models over conventional prediction models.

Although the better predictive performance of XGBoost has been demonstrated, some limitations of this study must be highlighted. This study's retrospective design was a major limitation, preventing the exclusion of all potential biases. In addition, patients were recruited from a single-center, which reduced its generalizability. Finally, the lack of external verification limits its clinical application. Hence, further studies should be implemented in larger cohorts with external validation.

To conclude, we developed a prediction model using XGBoost, a machine learning algorithm, to predict the live birth rate of IVF/ICSI patients based on their clinical characteristics, sex hormone levels, and COH features. Compared with the model constructed using conventional multivariate logistic regression, the XGBoost model showed a higher discriminative ability and net benefit, indicating its potential clinical value.

### Conflicts of interest

None.

### References

- De Geyter C. Assisted reproductive technology: impact on society and need for surveillance. *Best Pract Res Clin Endocrinol Metab* 2019;33:3–8. doi: 10.1016/j.beem.2019.01.004.
- Schirmer DA, Kulkarni AD, Zhang Y, Kawwass JF, Boulet SL, Kissin DM. Ovarian hyperstimulation syndrome after assisted reproductive technologies: trends, predictors, and pregnancy outcomes. *Fertil Steril* 2020;114:567–578. doi: 10.1016/j.fertnstert.2020.04.004.
- Devesa M, Martínez F, Coroleu B, Rodríguez I, González C, Barri PN. Ovarian response to controlled ovarian hyperstimulation in women with cancer is as expected according to an age-specific nomogram. *J Assist Reprod Genet* 2014;31:583–588. doi: 10.1007/s10815-014-0183-6.
- Hamdine O, Eijkemans MJC, Lentjes EGW, Torrance HL, Macklon NS, Fauser B, *et al*. Antimüllerian hormone: prediction of cumulative live birth in gonadotropin-releasing hormone antagonist treatment for in vitro fertilization. *Fertil Steril* 2015;104. 891.e2-898.e2. doi: 10.1016/j.fertnstert.2015.06.030.
- Portal A, Sunyach C, Loundou A, Lacroix-Paulmye O, Perrin J, Courbiere B. Nomograms for predicting adverse obstetric outcome in IVF pregnancy: a preliminary study. *Birth* 2021;48:186–193. doi: 10.1111/birt.12528.
- Li W, Yin Y, Quan X, Zhang H. Gene expression value prediction based on XGBoost algorithm. *Front Genet* 2019;10:1077. doi: 10.3389/fgene.2019.01077.
- Shi X, Wong YD, Li MZ, Palanisamy C, Chai C. A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accid Anal Prev* 2019;129:170–179. doi: 10.1016/j.aap.2019.05.005.
- Liu P, Fu B, Yang SX, Deng L, Zhong X, Zheng H. Optimizing survival analysis of XGBoost for ties to predict disease progression of breast cancer. *IEEE Trans Biomed Eng* 2021;68:148–160. doi: 10.1109/TBME.2020.2993278.
- Brykov MN, Petryshynets I, Pruncu CI, Efremenko VG, Pimenov DY, Giasin K, *et al*. Machine learning modelling and feature engineering in seismology experiment. *Sensors (Basel)* 2020;20:4228. doi: 10.3390/s20154228.
- Chamberlain JM, Chamberlain DB, Zorc JJ. Machine learning and clinical prediction rules: a perfect match? *Pediatrics* 2020;146: e2020012203. doi: 10.1542/peds.2020-012203.
- McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113,873 women. *BMJ* 2016;355:i5735. doi: 10.1136/bmj.i5735.
- McLernon DJ, Steyerberg EW, Te Velde ER, Lee AJ, Bhattacharya S. An improvement in the method used to assess discriminatory ability when predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation. *BMJ* 2018;362:k3598. doi: 10.1136/bmj.k3598.
- Lu H, Ma X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* 2020;249:126169. doi: 10.1016/j.chemosphere.2020.126169.
- Yuan KC, Tsai LW, Lee KH, Cheng YW, Hsu SC, Lo YS, *et al*. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform* 2020;141:104176. doi: 10.1016/j.ijmedinf.2020.104176.
- Gu SC, Zhou J, Yuan CX, Ye Q. Personalized prediction of depression in patients with newly diagnosed Parkinson's disease: a prospective cohort study. *J Affect Disord* 2020;268:118–126. doi: 10.1016/j.jad.2020.02.046.
- 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;17:317. doi: 10.1186/s12967-019-2062-5.
- Amini P, Ramezani F, Parchehbaf-Kashani M, Maroufzadeh 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;15:128–134. doi: 10.22074/IJFS.2020.134582.

**How to cite this article:** Chen Z, Zhang D, Zhen J, Sun Z, Yu Q. Predicting cumulative live birth rate for patients undergoing *in vitro* fertilization (IVF)/intracytoplasmic sperm injection (ICSI) for tubal and male infertility: a machine learning approach using XGBoost. *Chin Med J* 2022;135:997–999. doi: 10.1097/CM9.0000000000001874