PREDICTING THE OUTCOMES OF *IN VITRO* FERTILIZATION PROGRAMS USING A RANDOM FOREST MACHINE LEARNING MODEL

Vladimirsky GM[1], Zhuravleva MA[1], Dashieva AE[2] ✉, Korneeva IE[2], Nazarenko TA[2]

[1] Higher School of Economics National Research University, Moscow, Russia

[2] Kulakov National Medical Scientific Centre for Obstetrics, Gynecology and Perinatal Medicine, Moscow, Russia

Currently, *in vitro* fertilization (IVF) with embryo transfer is the main method of treatment of all forms of infertility, but successful cases ending in pregnancy still account for only a third of all cycles performed. It is necessary to take into account many parameters and investigate the connections between them in order to properly evaluate the results of IVF. Over the past decades, a number of IVF prediction models have been developed with the aim at assessing the outcomes in individual cases, but, given the generally poor prognostic capacity, only a few of them have proven to be clinically significant. This study aimed to create nonlinear IVF outcomes prediction models and identify the most significant factors affecting the said outcomes. Using a database containing information on more than 700 indicators of 7004 women aged 18 to 45 years who participated in the IVF program in Russia from 2010 to 2020, we trained a random forest model that predicted pregnancy in the IVF cycle with ROC-AUC = 0.69. This paper describes 20 most successful predictors of the resulting model and interprets their contribution to the prognosis. Of these, body mass index (BMI) and the number of received and fertilized oocytes have been covered in the scientific literature previously as predictors of IVF outcomes, but other parameters, such as anamnestic data, previous participation in an IVF program (number of cases and their results), as well as serum concentration of AMH, rarely appear in foreign prognostic models.

Keywords: IVF, infertility, prognostic model, random forest

ПРОГНОЗИРОВАНИЕ ИСХОДОВ ПРОГРАММ ЭКСТРАКОРПОРАЛЬНОГО ОПЛОДОТВОРЕНИЯ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ «СЛУЧАЙНЫЙ ЛЕС»

Г. М. Владимирский[1], М. А. Журавлева[1], А. Э. Дашиева[2] ✉, И. Е. Корнеева[2], Т. А. Назаренко[2]

[1] Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

[2] Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии имени В. И. Кулакова, Москва, Россия

Программы экстракорпорального оплодотворения (ЭКО) с переносом эмбрионов в настоящее время являются основными методами лечения всех форм бесплодия, однако успешные случаи, заканчивающиеся наступлением беременности, по-прежнему составляют лишь треть из числа всех проведенных циклов. Для оценки результатов лечения бесплодия методом ЭКО необходимо учитывать множество показателей и изучать взаимосвязь между ними. За последние десятилетия был разработан ряд моделей прогнозирования ЭКО для оценки индивидуальных результатов лечения, но лишь немногие из них оказались клинически значимыми, из-за их плохой прогностической способности. Целью исследования было создать нелинейные модели прогнозирования исходов ЭКО и выявить наиболее значимые факторы, влияющие на результат лечения. На основе базы данных, содержащей информацию по более чем 700 показателям 7004 женщин в возрасте от 18 лет до 45 лет, прошедших лечение в программе ЭКО на территории Российской Федерации с 2010 по 2020 г., была обучена модель «Случайного леса», прогнозирующая беременность в цикле ЭКО с ROC-AUC = 0,69. Описаны 20 самых успешных предикторов полученной модели с интерпретацией их вклада в прогнозирование. Из них индекс массы тела (ИМТ), число полученных и оплодотворившихся ооцитов описаны в научной литературе как предикторы исхода программ ЭКО, в то время как другие признаки, например анамнестические данные, количество и исходы предыдущих программ ЭКО, а также сывороточная концентрация АМГ, редко фигурируют в зарубежных прогностических моделях.

Ключевые слова: ЭКО, бесплодие, прогностическая модель, «Случайный лес»

Infertility is a problem affecting tens of millions of families. The development of assisted reproductive technologies (ART) gives such couples the hope that they can be parents. According to the report by Russian Association of Human Reproduction, in 2020, 148660 ART cycles were performed in the Russian Federation (RF), and about 34250 children were born. However, despite the population's need for this treatment being satisfied, clinical pregnancy occurs only in 34.8% of all embryo transfers [1]. There are many factors influencing the outcome of IVF, and they complicate the assessment of effectiveness of the cycles.

Therefore, development of a decision-making tool based on the analysis of these factors could improve the quality of medical care and counseling for patients in an IVF program.

Scientific literature offers several machine learning models that predict IVF outcomes and help identify women's characteristics and parts of the the program's protocol that affect such prediction the most [2].

Linear models are the most common approach to predicting results of IVF. A 2020 review identified 35 such models, all of them based on either logistic regression or Cox regression [3].

Often, such studies do not assess the quality of the models, although there are preferred methods for this, like ROC-AUC and c-statistics. ROC (receiver operating characteristic) analysis and the resulting ROC curve underpin qualitative assessment of predictive models. ROC analysis implies building four-field tables and measuring the model's sensitivity and specificity. ROC curve is a graphical plot that allows evaluating quality of the model by two classes. The ordinate axis is frequency of true positive results (sensitivity), and the abscissa axis is frequency of false positive results (specificity). The values are from 0 to 1 (that is, from 0 to 100%). The resulting curve shows the dependence of correctly classified positive cases on the number of incorrectly classified negative cases. In an ideal classifier, the ROC curve graph passes through the upper-left corner, where the proportion of true positive cases is 1.0, or 100% (ideal sensitivity), and the proportion of false positive cases is 0. Another characteristic used in assessment of quality of the model is area under curve (AUC). The higher the AUC, the higher predictive power of the model. More often, AUC is intended for comparative analysis of several models. In literature, the values of ROC-AUC in IVF results prediction range from 0.58 to 0.73 [3–12].

Typically, linear IVF success predictive models include about seven attributes. The most common are the woman's age, causes of infertility, outcome of previous pregnancies and IVF program enrollments, number of oocytes obtained and embryos transferred [4, 5, 9–11]. Some researchers believe that a limited number of attributes, which are used in the vast majority of studies, makes the predictive power of models rather modest, and advocate identification of new factors affecting outcome of the procedure [13].

Despite being commonly used, logistic regression models have a number of disadvantages. For example, several studies have revealed the nonlinear character of relationship between the success of IVF and key attributes, such as woman's age, number of oocytes obtained, and treatment initiation year [10, 11]. In such cases, cubic spline function can enable data interpolation (for example, age) and make linear models non-linear, or the data can be transformed to be polynomial [8, 10, 11]. Still, such modifications of linear models are based on a simple (polynomial) relationship between the target variable and the attributes.

Besides, logistic regression models are interpreted, and they do not possess high predictive power. Therefore, many researchers have turned to non-linear, non-interpretable machine learning models relying on random forest, gradient boosting, and neural networks. Random forest and gradient boosting are often considered the most advanced methods applicable to binary classification problems involving tabular data, since they tend to be unequaled in accuracy and generalization power [2]. As a rule, ROC AUC for such models ranges from 0.68 to 0.86, which is higher than that for linear classifiers [14–16].

The main limitation of non-linear non-interpreted models is the complexity of estimation of significance of each attribute to the prediction model. However, methods developed in the recent years enable interpretation of attributes for any machine learning model, regardless of their complexity. For this study, we used the SHAP method, which is based on the Shapley value, a concept from cooperative game theory. This method calculates contribution of each attribute to the prediction relying on the approximated Shapley value (average contribution of an attribute to all coalitions thereof) [17], which allows accurate predictions of the IVF programs outcomes.

This study aimed to build nonlinear IVF outcome prediction models and identify the most significant factors affecting the said outcome.

METHODS

Clinical material

To build the model, we used data covering the characteristics of 7004 women and presenting the outcomes of their participation in the IVF programs. They were treated at 17 ART clinics in RF from 2011 to 2020. The inclusion criteria were age from 18 to 45 years, and infertility for any reason (N97). The exclusion criteria were contraindications for ART and pregnancy, as per the Order of the Ministry of Health of the Russian Federation #803n of July 31, 2020"On the procedure of application of assisted reproductive technologies, respective contraindications and restrictions."

Figure 1 shows the distribution of clinics participating in the study by subjects of the RF. For the purpose of collection of the material, we developed questionnaires listing 770 attributes, which were filled by specialists at the said clinics. The resulting data were broken into several blocks: social characteristics of patients (124 questions), medical history (171 questions), which included data on the state of somatic health (58 questions), gynecological health (108 questions), history of infertility and treatment methods (73 questions), laboratory examination data (6 points), data on the patient's partner (210 questions), data on the protocol of ovarian stimulation (7 questions) and embryological stage (30 questions), support for the luteal phase, and outcome of participation in the IVF program.

Data processing and analysis

Preprocessing of the data for the model included selection of the minimum value among several analyses of serum hormone levels (anti-mullerian hormone, or AMH; follicle stimulating hormone, or FSH; lutenizing hormone, or LH; thyroid stimulating hormone, TSH; prolactin). After removal of sparse and duplicate data, there remained 408 attributes. Gaps were filled with averages. We used odds ratio (OR) for the statistical analysis [18]; the respective p value was calculated as per [19].

Selection of attributes and interpretation of their significance

In this study, we used random forest, a machine learning method that relies on an ensemble of decision trees for classification tasks. Each individual tree in such a forest gives a prediction of a class, and the class with the highest number of votes becomes the prediction. The purpose of this work was to forecast pregnancy after IVF.

After building the random forest, we applied the Gini coefficient to measure inequality of the attributes. This coefficient allows comparing distribution of an attribute in a sample with a different number of units [20]. The model used for attribute allocation had hyperparameters, which are manually adjusted before training and allow maximization of the ROC AUC value in a five-fold cross validation; subsequently the model was trained on a full dataset. Selecting the attributes in the optimal amount, we applied the recursive selection method with five-fold cross validation, which implies removal of the least significant attribute at each step. All of the above methods were used in the implementation of the scikit-learn library [21]. The SHAP method [17], designed for interpretation of significance of attributes in a non-linear model, enabled extended interpretation of the results.

Models used

Because of the large number of binary and categorical attributes, as well as nonlinear dependencies between the attributes and the

Fig. 1. Geographical scope of the study

target variable, we used the random forest model implemented in the scikit-learn library [21]. GridSearch method with five-fold cross validation [21] enabled selection of the model's parameters, and the classes of the model were assessed using ROC AUC, which is less sensitive to the imbalance of classes in the data. Ultimately, the best parameters for the random forest model were the maximum depth of 50, not less than 2 objects per sheet, and 2000 trees in total. In addition, we tested the Catboost classifier model [22], which was chosen for its built-in support of categorical attributes that distinguishes it from other implementations of the gradient boosting algorithm. The target variable for all trained models was pregnancy (or lack thereof).

RESULTS

Recursive selection of attributes has shown that ROC AUC reaches its maximum (0.69) when training of the random forest involves 220 of them. In cross validation, the maximum Catboost

ROC AUC value was 0.68, therefore, further on, we used the random forest model, which is more convenient for interpretation. Figure 2 presents dynamics of this model's ROC AUC metric when the attributes are gradually removed therefrom.

At the outset, gradual removal of the attributes translates into insignificant changes of the ROC AUC value, which drops abruptly only when the their number goes below 33. Therefore, we chose 33 as the optimal amount of attributes in the model, with the ROC AUC value therewith reaching 0.69.

Gini coefficient was applied to establish the significance of 20 attributes with the greatest impact on the prediction (Fig. 3).

The attributes most significant for the prediction were date of birth (age), the number of fertilized oocytes, and the total number thereof, which is consistent with the data reported by international studies [16].

Compared to other hormone indicators, serum AMH had the greatest weight in prediction, but in international studies,
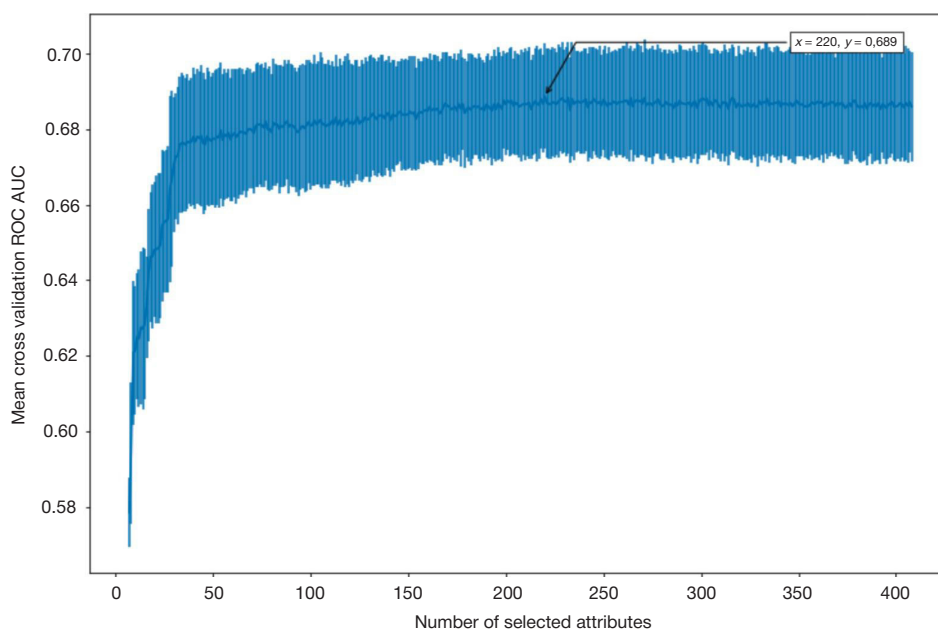

Fig. 2. Mean cross validation ROC AUC dynamics during elimination of attributes. The diagram shows the maximum ROC AUC and the standard deviation of the value in cross validation
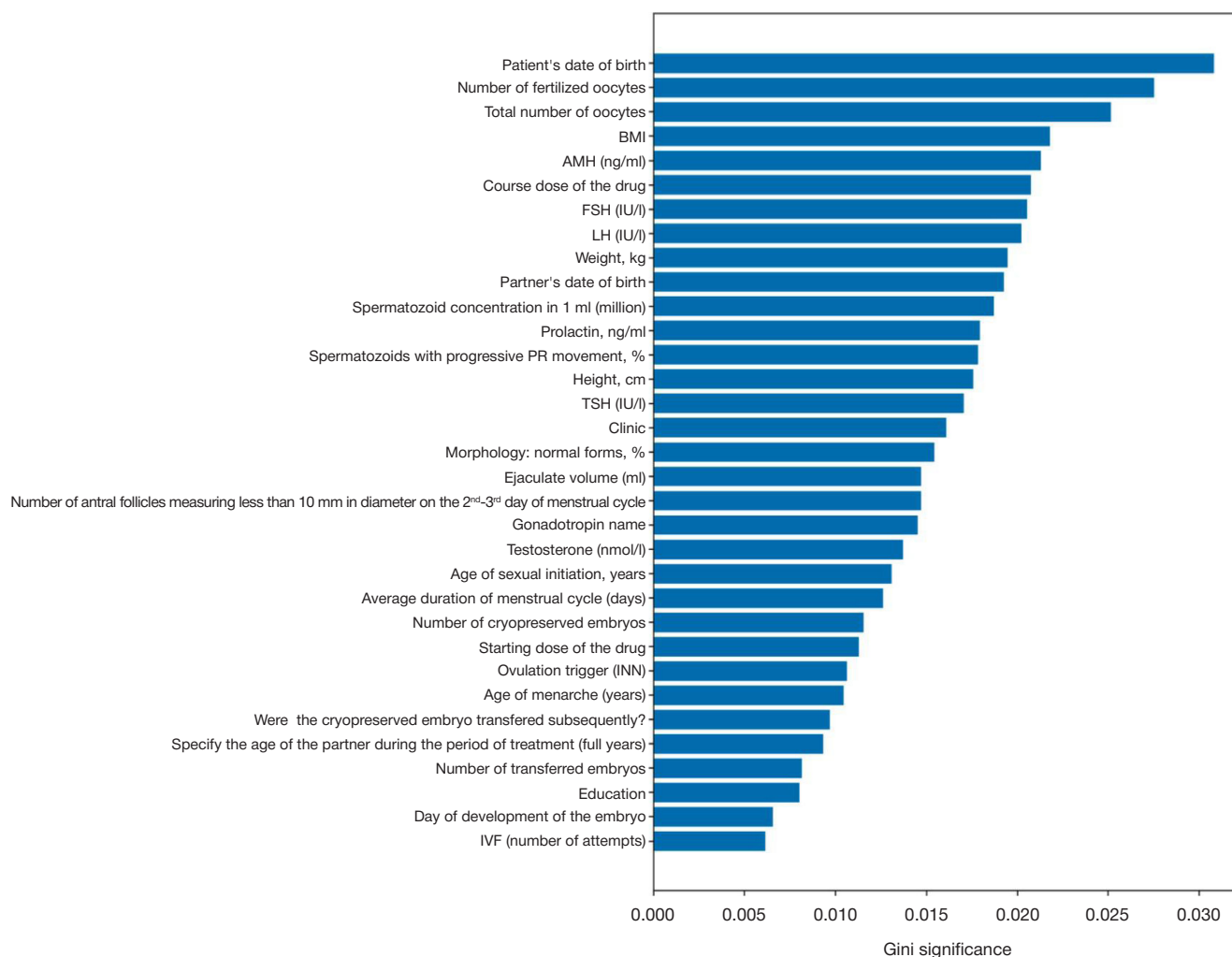
Fig. 3. Gini significance for 33 best attributes learned after training the random forest on 220 attributes

it is much less common than the levels of gonadotropins (FSH and LH) [8]. The clinic where the patient underwent IVF was also a significant factor.

The top 20 attributes were analyzed additionally using the SHAP method. As shown on Fig. 4, the chances of a successful outcome of IVF, as predicted by the model, grow along with the values of such attributes as the number of fertilized oocytes, date of birth of the patient, the level of AMH, use of progesterone in the luteal phase of the cycle.

We have built models predicting the outcome of IVF for individual infertility diagnoses. The resulting ROC AUC values did not exceed the value of the metric for the entire sample, which allows concluding that using models for certain types of infertility is impractical (Table).

Women with unsuccessful IVF in the past have a lower probability of a successful IVF than women joining the program for the first time or whose previous attempts were successful (OR = 0.7675; $p < 0.0001$). Therefore, it is natural that the variable reflecting the number of past attempts is one of the significant attributes selected via the random forest classifier, which performed best in the cross validation trial. Nevertheless, we consider the model described by us to be more relevant for the Russian population than the foreign models described in the literature.

DISCUSSION

Our data indicate that prognostic quality of the current random forest model (ROC AUC = 0.69) is comparable to that of the

similar models described in foreign studies. For example, a recent report presented a model with the best ROC AUC value of 0.68 [14].

Despite the said comparability of ROC AUC of our model and foreign models, in most cases, they are based on different criteria of selection of pairs, even with similar target variable. Some models described in the foreign literature disregard data on the past IVF attempts and consider the outcomes of the first program a woman participates in [14]. Our IVF model relies on the results of previous cycles: 40.9% of the women whose histories comprised the training dataset had unsuccessful IVF attempts previously. Thus, our model allows predicting IVF outcomes for a single cycle, which is an advantage, since some models by foreign researchers prognosticate cumulative success for several IVF cycles [10, 11].

All the above factors make comparison of the models by numerical indicators only partially objective. For example, it can be assumed that our model has a higher ROC AUC than the earlier described model [14], since it factors in data on the previous IVF attempts, which, for the 40.9% of women who enjoyed no success before (and whose data was part of the training dataset), translates into 92.95% chance of failure in the next IVF cycles.

Considering the effectiveness of the developed model from a clinical perspective, we should note the identified most important IVF outcome predictors peculiar to Russian infertile couples. The list of such attributes includes both those traditionally accounted for (woman's age, number of fertilized
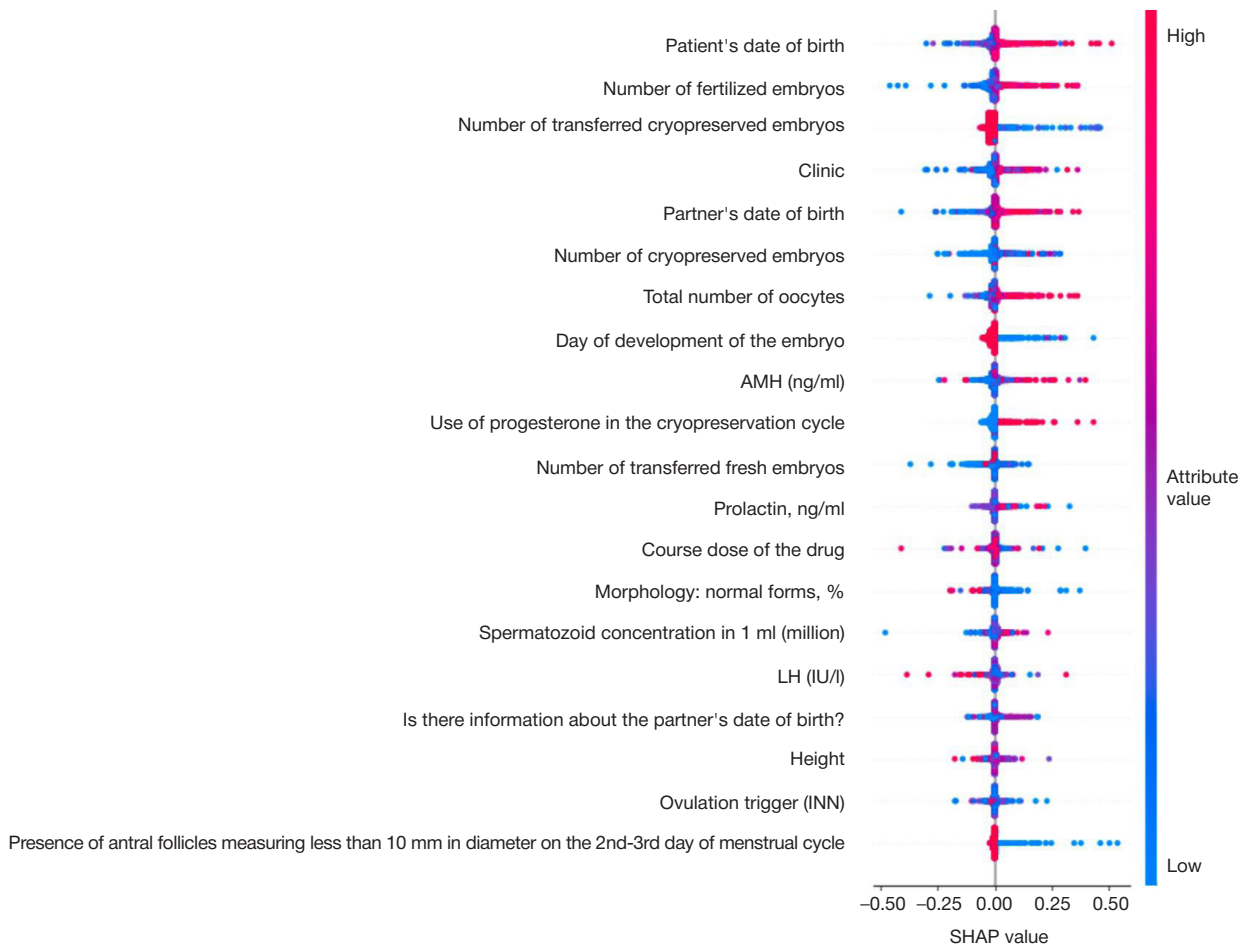
Fig. 4. Diagram of SHAP values for the 20 most significant attributes. A positive class means successful IVF, negative class — failure thereof

oocytes, total number of oocytes, BMI, AMH level) and the predictors typically unregarded. For example, progesterone drugs during the luteal phase were shown to be associated with successful outcomes. Although prescribing progesterone is a routine clinical tactic, until now there has been no mathematically proven justification for the need to support the luteal phase of the induced cycle. Besides, there is now an objective confirmation of the negative effect previous unsuccessful IVF attempts have on the planned one. This fact, apparently, necessitates a review of the treatment tactics when the patient's history cites several (four, in this study) IVF failures. Despite the fact that today IVF is a routine infertility treatment method, and it would seem that all clinics apply standard protocols and technology, our model revealed establishment-dependent differences based on the data provided by them, which may justify an analysis of their approaches. An interesting fact uncovered in this study is the lack of dependence of IVF outcome on the confirmed infertility diagnosis in situations when all other significant factors are similar. This is contradictory to the results of many studies that

seek a link between success/failure of IVF and infertility as a nosology.

CONCLUSIONS

Over the past decades, a number of IVF prediction models have been developed that aim at assessing the outcomes in individual cases, but, due to the insufficient prognostic capacity and statistical methods used, only a few of them have proven to be clinically significant. Machine learning, which enables interpretation of data and development of predictive models, finds increasingly wider application in clinical practice, especially for complex systems with multiple variables. In this study, we have built a model that predicts the outcome of IVF cycles with satisfactory forecasting efficiency, identified the important factors of IVF effectiveness, and uncovered interactions between them. We will continue to explore practical applications of the model seeking to assess the impact of variables on the efficacy of treatment.

Table. Results of five-fold cross validation of the random forest model, individual infertility diagnoses, in the context of selection of hyperparameters using the GridSearch method

Infertility diagnosis (ICD-10)	Number of patients	ROC-AUC
Female infertility associated with anovulation (N 97.0)	1279	0.69 ± 0.02
Female infertility of tubal origin (N 97.1)	3191	0.66 ± 0.02
Female infertility of uterine origin (N 97.2)	1248	0.66 ± 0.05
Female infertility associated with male factors (N 97.4)	3226	0.68 ± 0.01

References

1. Korsak VS, Smirnova AA, Shurygina OV. Registr VRT Obshherossijskoj obshhestvennoj organizacii «Rossijskaja associacija reprodukcii cheloveka». Otchet za 2020 god. Problemy reprodukcii. 2022; 28 (6): 12–27. DOI: 10.17116/repro20222806112. Russian.
2. Hassan MR, Al-Insaif S, Hossain MI, Kamruzzaman J. A machine learning approach for prediction of pregnancy outcome following IVF treatment. Neural Computing and Applications. 2018; 32: 2283–97.
3. Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. Human Reproduction. 2020; 35: 100–116.
4. Lintsen A, Eijkemans M, Hunault C, Bouwmans C, Hakkaart L, Habbema J, et al. Predicting ongoing pregnancy chances after IVF and ICSI: A national prospective study. Human Reproduction. 2007; 22: 2455–62.
5. van Weert JM, Repping S, van der Steeg JW, Steures P, van der Veen F, Mol BW. A prediction model for ongoing pregnancy after in vitro fertilization in couples with male subfertility. J Reprod Med. 2008; 53 (4): 250–6. PMID: 18472647.
6. Jones CA, Christensen AL, Salihu H, Carpenter W, Petrozzino J, Abrams E, et al. Prediction of individual probabilities of livebirth and multiple birth events following in vitro fertilization (IVF): A new outcomes counselling tool for IVF providers and patients using HFEA metrics. Journal of experimental clinical assisted reproduction. 2011; 8: 3.
7. Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: A prospective study of 144,018 treatment cycles. PLoS Medicine. 2011; 8: e1000386.
8. van Loendersloot LL, van Wely M, Repping S, Bossuyt PMM, van der Veen F. Individualized decision-making in IVF: Calculating the chances of pregnancy. Human Reproduction. 2013; 28: 2972–80.
9. Dhillon R, McLernon D, Smith P, Fishel S, Dowell K, Deeks J, et al. Predicting the chance of live birth for women undergoing IVF: A novel pretreatment counselling tool. Human Reproduction. 2015; 31: 84–92.
10. 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; p. i5735.
11. McLernon DJ, Raja EA, Toner JP, Baker VL, Doody KJ, Seifer DB. Predicting personalized cumulative live birth following in vitro fertilization. Fertility and Sterility. 2022; 117 (2): 326–38.
12. Vaegter KK, Lakic TG, Olovsson M, Berglund L, Brodin T, Holte J. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single- embryo transfers. Fertility and Sterility. 2017; 107 (3): 641–48.
13. Ottosen LDM, Kesmodel U, Hindkjaer J, Ingerslev HJ. Pregnancy prediction models and eSET criteria for IVF patients–do we need more information? Journal of assisted reproduction and genetics. 2007; 24 (1): 29–36.
14. Xu T, d. F. Veiga A, Hammer KC, Paschalidis ICh, Mahalingaiah Sh. Informative predictors of pregnancy after first IVF cycle using eIVF practice highway electronic health records. Scientific Reports. 2022; 12: 1–9.
15. C N Barreto N, Castro GZ, Pereira RG, Pereira FAN, Reis FM, Junior WM, et al. Predicting in vitro fertilization success in the Brazilian public health system: a machine learning approach. Med Biol Eng Comput. 2022; 60 (7): 1851–61. DOI: 10.1007/s11517-022-02569-1. Epub 2022 May 4. PMID: 35508786.
16. Barnett-Itzhaki Z, Elbaz M, Butterman R, Amar D, Amitay M, Racowsky C, et al. Machine learning vs. classic statistics for the prediction of IVF outcomes. Journal of Assisted Reproduction and Genetics. 2020; 37: 2405–12.
17. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Available from: https://arxiv.org/abs/1705.07874, may 22 2017.
18. Morris JA, Gardner MJ. Statistics in Medicine: Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. BMJ. 1988; 296: 1313–6.
19. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC, 2003.
20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge, 2017.
21. Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011; 12: 2825–30.
22. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. Catboost: Unbiased boosting with categorical features. 2017. Available from: https://arxiv.org/abs/1706.09516.

Литература

1. Корсак В. С., Смирнова А. А., Шурыгина О. В. Регистр ВРТ Общероссийской общественной организации «Российская ассоциация репродукции человека». Отчет за 2020 год. Проблемы репродукции. 2022; 28 (6): 12–27. DOI: 10.17116/repro20222806112.
2. Hassan MR, Al-Insaif S, Hossain MI, Kamruzzaman J. A machine learning approach for prediction of pregnancy outcome following IVF treatment. Neural Computing and Applications. 2018; 32: 2283–97.
3. Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. Human Reproduction. 2020; 35: 100–116.
4. Lintsen A, Eijkemans M, Hunault C, Bouwmans C, Hakkaart L, Habbema J, et al. Predicting ongoing pregnancy chances after IVF and ICSI: A national prospective study. Human Reproduction. 2007; 22: 2455–62.
5. van Weert JM, Repping S, van der Steeg JW, Steures P, van der Veen F, Mol BW. A prediction model for ongoing pregnancy after in vitro fertilization in couples with male subfertility. J Reprod Med. 2008; 53 (4): 250–6. PMID: 18472647.
6. Jones CA, Christensen AL, Salihu H, Carpenter W, Petrozzino J, Abrams E, et al. Prediction of individual probabilities of livebirth and multiple birth events following in vitro fertilization (IVF): A new outcomes counselling tool for IVF providers and patients using HFEA metrics. Journal of experimental clinical assisted reproduction. 2011; 8: 3.
7. Nelson SM, Lawlor DA. Predicting live birth, preterm delivery, and low birth weight in infants born from in vitro fertilisation: A prospective study of 144,018 treatment cycles. PLoS Medicine. 2011; 8: e1000386.
8. van Loendersloot LL, van Wely M, Repping S, Bossuyt PMM, van der Veen F. Individualized decision-making in IVF: Calculating the chances of pregnancy. Human Reproduction. 2013; 28: 2972–80
9. Dhillon R, McLernon D, Smith P, Fishel S, Dowell K, Deeks J, et al. Predicting the chance of live birth for women undergoing IVF: A novel pretreatment counselling tool. Human Reproduction. 2015; 31: 84–92.
10. 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; p. i5735.
11. McLernon DJ, Raja EA, Toner JP, Baker VL, Doody KJ, Seifer DB. Predicting personalized cumulative live birth following in vitro fertilization. Fertility and Sterility. 2022; 117 (2): 326–38.
12. Vaegter KK, Lakic TG, Olovsson M, Berglund L, Brodin T, Holte J. Which factors are most predictive for live birth after in vitro fertilization and intracytoplasmic sperm injection (IVF/ICSI) treatments? Analysis of 100 prospectively recorded variables in 8,400 IVF/ICSI single-embryo transfers. Fertility and Sterility. 2017; 107 (3): 641–48.
13. Ottosen LDM, Kesmodel U, Hindkjaer J, Ingerslev HJ. Pregnancy prediction models and eSET criteria for IVF patients–do we need more information? Journal of assisted reproduction and genetics.

2007; 24 (1): 29–36.

14. Xu T, d. F. Veiga A, Hammer KC, Paschalidis ICh, Mahalingaiah Sh. Informative predictors of pregnancy after first IVF cycle using eIVF practice highway electronic health records. Scientific Reports. 2022; 12: 1–9.

15. C N Barreto N, Castro GZ, Pereira RG, Pereira FAN, Reis FM, Junior WM, et al. Predicting in vitro fertilization success in the Brazilian public health system: a machine learning approach. Med Biol Eng Comput. 2022; 60 (7): 1851–61. DOI: 10.1007/s11517-022-02569-1. Epub 2022 May 4. PMID: 35508786.

16. Barnett-Itzhaki Z, Elbaz M, Butterman R, Amar D, Amitay M, Racowsky C, et al. Machine learning vs. classic statistics for the prediction of IVF outcomes. Journal of Assisted Reproduction and Genetics. 2020; 37: 2405–12.

17. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Available from: https://arxiv.org/abs/1705.07874, may 22 2017.

18. Morris JA, Gardner MJ. Statistics in Medicine: Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. BMJ. 1988; 296: 1313–6.

19. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC, 2003.

20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge, 2017.

21. Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011; 12: 2825–30.

22. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. Catboost: Unbiased boosting with categorical features. 2017. Available from: https://arxiv.org/abs/1706.09516.