Prediction of Clinical Pregnancy Occurrence After ICSI using Decision Tree and Support Vector Machine Methods

Mahdieh Kafaee Ghaeini[1] ID, Mohammad Reza Amin-Naseri[2*] ID, Marzieh Aghahoseini[3] ID

1. MSc Student of Industrial Engineering-System Management, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran
2. Professor, Industrial and Systems Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran
3. Professor of Medical School, Tehran University of Medical Sciences, Tehran, Iran

Article Info

Corresponding Information:
Mohammad Reza Amin-Naseri, Professor, Industrial and Systems Engineering, Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran
Email: amin_nas@modares.ac.ir
Tel: 021-82883344

ABSTRACT

Background & Objective: Studies have shown that despite the numerous research carried out regarding infertility treatment, there is still a long way to treat this disease satisfactorily. Spending a lot of time and money on infertility treatments proves the necessity of designing a model which could predict the result of treatment methods with an acceptable accuracy; a model that could help physicians to get rid of trial and error for treatment methods which should step by step be applied on an infertile couple. Intracytoplasmic Sperm Injection (ICSI) is one of the assisted reproductive techniques. Statistics have indicated that the probability of pregnancy occurrence is only about 30% using this method. In this paper, a model which could predict the result of (ICSI) was presented using the decision tree and support vector machine methods.

Materials & Methods: The applied data were collected in seven months from December 2012 to June 2013 by analyzing 251 treatment cycles in Omid Fertility Clinic. Input variables of the model were parameters like couple's medical records, hormonal tests, the cause of infertility, and the like. The output variable was the occurrence or nonoccurrence of the clinical pregnancy (the pregnancy resulting in the formation of the fetal heart). One of the innovations of this study was that the input variables of the model were only preoperative, while in previous studies, having information about some of the surgery stages, such as quality of the egg and the like, was required to anticipate the result of the surgery.

Results: The obtained accuracy using the decision tree and support vector machine methods were 70.3% and 75.7%, respectively.

Conclusion: The results of the current study demonstrated that the support vector machine method had a better performance compared to the decision tree method. Presented model predicts the occurrence or nonoccurrence of a clinical pregnancy follows (ICSI), with a precision of 75.7%.

Keywords: Infertility, Intracytoplasmic sperm injection, Decision tree, Support vector machine

Introduction

Infertility refers to couples' inability to have a child a year after having regular unprotected sexual intercourses (1). The prevalence of infertility in the world, especially in traditional societies like Iran, has caused many social, economic, and family problems for infertile couples and society. Based on a report released by the World Health Organization, this disease has affected more than 80 million people worldwide (2).

Due to a cause or causes of infertility, a physician may step-by-step use different therapies for an infertile couple; if a pregnancy does not occur at each stage, another treatment method which is in the next priority is used.

That is why the process of infertility treatment is often time-consuming and imposes a lot of economic costs on society and infertile couples. Therefore, it is useful to design a model that can predict the results of the treatment method with acceptable accuracy. By anticipating the outcome of the treatment method, such model aids physicians to choose a treatment method at a lower risk. It also reduces the probability of wasting a plethora of time and money spent on the treatment.

In this paper, using some characteristics of infertile couples, the occurrence or nonoccurrence of a clinical pregnancy was predicted following the implementation of

Intracytoplasmic Sperm Injection (ICSI). These characteristics include medical records of a husband and a wife, hormonal tests, the cause of infertility, and the like. Reviewing previously carried out studies has shown that various studies have yielded different and even contradictory results in identifying some parameters affecting the failure or success of treatment methods. Some studies also suggest that the low accuracy in prediction is due to the lack of comprehensive application of the parameters affecting the treatment method. Therefore, there is still a need for conducting studies to examine the issue.

On the other hand, in previous studies, having information on some stages of the surgery, such as the quality of eggs taken from the ovary, the quality of embryo fertilization and the like, was necessary to predict the outcome of the surgery. The advantages of this study are that the input variables in the proposed model are all preoperative. Accordingly, a physician can treat a patient with an acceptable percentage error without a need for trial and error this treatment method. If it is predicted that the surgery results in the nonoccurrence of a pregnancy with the formation of a fetal heart, the physician can suggest another treatment option which is the next priority considering the couples' condition.

ICSI Technique in the Treatment of Infertility

The process of infertility treatment is sometimes a long process, and after examining the primary methods of treatment, if the pregnancy does not occur, a physician will suggest the use of the assisted reproductive technology (ART). The assisted reproductive technology has different types, such as IUI, IVF, ICSI, GIFT, and ZIFT. Considering a couple's condition like a cause or causes of infertility, a physician suggests one of the mentioned methods.

ICSI is an invasive technique for infertility treatment. In this method, the egg plasma membrane is penetrated, and the sperm is injected into it (3). Two to three days later, some embryos, according to their quality and considering the conditions of the patient, are transmitted to the woman's uterus based on the physician's opinion.

Follow-up Steps After the Surgery

There are three stages after using the assisted reproductive techniques to diagnose the occurrence of pregnancy. In the first stage, a woman's blood test indicates whether the embryo has succeeded in the implantation in the mother's uterus or not. This stage which occurs two weeks after the embryo transfer to the woman's uterus is known as a chemical pregnancy. In the second stage, one week after the first stage, an ultrasound demonstrates whether the pregnancy sac is formed or not. In the third stage, one week after the second stage, an ultrasound, shows whether the embryo's heart is formed or not. This stage is known as a clinical pregnancy.

In some cycles, for unknown reasons, the woman's blood test or her pregnancy test is positive; however, the pregnancy sac is not formed. In some other cases, the pregnancy sac may form, but the fetus's heart does not form. In such cases, the physician has to conduct an abortion. Therefore, in the first few weeks after ICSI, the outcome of the third stage is of particular importance. That is why, in this paper, the outcome variables, i.e., the occurrence or nonoccurrence of the clinical pregnancy which is shown by the results of an ultrasound carried out five weeks after the embryo transfer, was considered.

The first pregnancy occurred using the assisted reproductive techniques was in 1978 (4). Since then, several studies have been done to identify the factors affecting these treatment methods. Some of these studies used statistical techniques to recognize these factors. One study has shown that in addition to the parameter of the number of embryos, other parameters, such as a woman's age, play key roles in determining the extent of pregnancy and multiple pregnancies. In this paper, it was mentioned that the results of some similar studies were different from the results of this study. For example, in a research conducted by Rawhon (2002), there was a significant relationship between women's younger age and the increase in pregnancy; however, there was no relationship between women's age and multiple pregnancies (5).

Another study carried out to investigate the relationship between women's BMI and the outcome of IVF or ICSI proved that pregnancy rates in obese women were significantly lower than other women (6). Other studies have confirmed that pregnancy rates among obese women were lower than those of normal-weight women. Additionally, among independent variables that affect predicting the birth of a newborn, the following variables are known: the age of a woman, her body mass index, and the age of a man (7). Another study showed that women's BMI did not affect the outcome of IVF and clinical pregnancy (8).

In addition to research on identifying factors affecting ART, many efforts have been undertaken to improve the process of infertility treatment and increase the probability of the occurrence of pregnancy. One of the sciences which aids medical sciences is the science of data mining. One of the studies conducted in this field is the application of a neural network to predict the results of IVF. The accuracy of predicting the results of this treatment method was 59% (9).

In another study, Bayesian networks were designed and used in decision support systems to choose the appropriate embryos for transfer in IVF (10). In another study, the decision tree method was combined with a genetic algorithm to predict the outcome of IVF. The accuracy, sensitivity, and specificity of the obtained model were 73.2%, 55.3%, and 85.2%, respectively (4). In another study, a Bayesian network model was used to predict pregnancy after IVF. In this study, AUC was the criterion for assessing the performance of the model. AUC_0, AUC_1, and AUC_2 (equivalent to non-occurrence of a pregnancy, a single pregnancy, and a twin pregnancy) were 74.1%, 67%, and 83.4%, respectively (2). Other studies have been conducted to predict the causes of couples' infertility. The results of

this study suggested that the support vector machine using a polynomial kernel function predicted with the highest accuracy (76.7%) **(11)**.

In another study, the authors used a diagnostic analysis method to predict pregnancy in the IVF treatment. The proposed model was able to predict the occurrence and nonoccurrence of a pregnancy with a precision of 51.22% and 74.07%, respectively. Therefore, this model is more suitable for predicting the negative results of the treatment (nonoccurrence of a pregnancy) **(12)**. Another study has demonstrated that applying the technique of analyzing the fundamental components of the data before the training process by the neural network can predict the performance of the model in anticipating the outcome of the IVF treatment slightly better than a model constructed solely by the neural network. Although the final results are still not satisfactory **(13)**. Another study carried out on 610 infants revealed that two data mining algorithms for analyzing the fundamental components and neural network together could extract almost all the available data **(14)**.

As mentioned earlier, to detect some of the parameters that affect the failure or success of the treatment methods, different researches have reached different and even contradictory results. Some studies have not succeeded because they could not fully apply the parameters affecting the treatment method in their data analyses. Therefore, carrying out studies in this field still seems necessary. Furthermore, in the previous studies, having information on some stages of the surgery, such as the quality of eggs taken from a mother's ovary, the quality of embryo fertilization, and the like, was essential to predict the outcome of the surgery. One of the advantages of this study is that the input variables in the proposed model are all preoperative. Therefore, if it is predicted that the surgery results in the nonoccurrence of a pregnancy, a physician can suggest another treatment option which is the next priority considering the couples' condition.

Materials and Methods

In the current paper, two data mining techniques were used for predicting the result of the ICSI treatment method. In the following, these techniques are briefly explained.

Decision Tree

As its name suggests, the structure of the decision tree is like a tree. Each node represents a test on an attribute, and each branch represents a possible value for that attribute. The leaves represent a label for each class. A classification procedure for each record begins with testing the introduced attribute at the root node. Then, it moves to a lower lever branch, based upon the obtained value for that attribute. This process continues by testing the next node on the selected branch. Finally, we reach a leaf which denotes the label of the considered class **(15)**.

For modeling data in this paper, the classification and regression tree (CART) type of decision tree was used. The

CART is a binary tree **(16)**, i.e., each node has only two branches. The CART works based on the Hunt algorithm **(17)**. According to this algorithm, the decision tree develops by recursively splitting training records to purer subsets (i.e., subsets of the same class, if possible) **(16)** until each class holds a label.

Among the most important features of this method, resistance to disorder[17] and outlying points, ability to solve nonlinear problems, plain interpretation of results and deriving if-then rules can be mentioned **(16)**.

Support Vector Machine

The support vector machine was developed for binary classification of data, assuming we have a problem with binary outputs. This method uses non-linear mappings to convert primary training data into higher dimensions. With new dimensions, the system seeks to find an optimal hyper plane which is a linear classifier and separates data of one class from the other **(16)**.

Among the features of this method, modeling complex nonlinear decision boundaries, high precision, finding the optimal solution for a problem **(16)**, and working with high-dimension inputs **(17)** can be mentioned.

Statistical Population and Research Variables

In this study, 330 ICSI cycles in Omid Fertility Clinic were analyzed from December 2012 to June 2013. After reviewing the literature and using the expertise and experience of a physician, nine features, which were the most important features affecting the outcome of the surgery, were considered as input variables. As mentioned earlier, all input variables were features determined before the treatment cycle. Additionally, the occurrence or nonoccurrence of a clinical pregnancy was considered as an output variable. Thus, for each treatment cycle, ten features were recorded and investigated. The features are briefly explained in Table 1. It is worth noting that the history of ART surgery and Anti-Mulirian Hormone were not examined in previous studies; however, they were included in the current research as suggested by the supervising physician. According to the table, three variables related to the causes of infertility **(6-8)** are clearly of a string type, and the other input variables are of a numeric type. Furthermore, the output variable is of binary type.

It should be noted data collected through medical records of patients and interview with them. Patients' consent to use their information in this research is considered and this information remains confidential.

Data Preprocessing

After collecting the data, those items, which were out of the scope of this study (e.g., cycles using frozen eggs or embryos, donated embryo, and cycles with unknown results for whatever reason), were excluded from preprocessing. At the end of this stage, 251 cycles remained. A proper code was also attributed to string variables; details of coding the variables are given in Table 1.

Table 1. The most important features affecting the result of ICSI (before the treatment cycle)

No.	Variables (unit)		Details	Abbr.	Input/Output
1	Wife's age (year)			F_age	
2	BMI (kg/m^2)		Calculated by dividing one's weight in kilogram by the square of one's height in meter; it is a measure that shows one's weight to height.	BMI	
3	History of ART surgery		The number of surgeries performed on the couple by ART methods (except IUI)	H_ART	
4	endometrial thickness (mm)		The endometrium is the inner lining of the uterus; the thicker the endometrium, the more suitable it is for pregnancy.	T_ando	
5	Duration of infertility (year)		The number of years that the couple did not bear a child without contraceptive methods.	D_inf	
6	Causes of infertility	Male factors	Normal male factors, severe male factors, TESE, and PESA are specified with code numbers 2, 3, 4, and 5, respectively (code 1 represents cycles whose infertility is due to a female or other factors; in other words, none of the mal factors is the cause of the cycle's infertility. The same process applies to female and other factors)	M_etio	Input
7		Female factors	Ovarian, uterine, and endometrial factors are specified with code number 2, 3, and 4. Cases identified with both ovarian and uterine factors are labeled with code 5, and cases identified with both ovarian and endometrial factors are labeled with code 6.	F_etio	
8		Other factors	Unknown and other factors are specified with code 2 and 3.	O_etio	
9	AntiMulirian Hormone		One of the women's hormonal tests which is the most critical hormonal test affecting her fertility, according to physicians.	AMH	
10	Clinical pregnancy		The ultrasound result about five weeks after transferring the embryo, a record belongs to class 1 if the heart is formed; otherwise, it belongs to class 0.	OP	Output

Missing Data

In some cases, the test result of one of the couples was not available in their records or, according to the doctor, performing some tests was not necessary. Also, for any reason, some information was not registered in some of these couples' records. Hence, there was some missing information in the collected data. To fill out the missing data in the input variables, except for the Anti-Mulirian Hormone, a median of each variable was considered instead of the missing value.

Results

Regarding the Anti-Mulirian Hormone variable, since the number of missing data was high, the use of the method mentioned above would lead to the dependence of this variable on the output variable, and the model would somehow be trained with cheating. Assume that the estimated values for the missing data of this variable were "A" in 0-labeled class and "B" in 1-labeled class. In predicting the output for experimental records, the Anti-Mulirian Hormones equal to "A" and "B" values would just be labeled with 0 and 1, respectively. As noted before, since the number of missing data for this variable was high, this could happen to several records and result in false accuracy of the model. Missing values of the Anti-Mulirian Hormone were predicted using a neural network to prevent the mentioned problem. According to the

physician, a wife's age and BMI are two critical variables in predicting the Anti-Mulirian Hormone, they were considered as input variables, and Anti-Mulirian Hormone's value was taken as the output of the model.

Modeling

In this section, the results of applying the decision tree and support vector machine methods on data were analyzed using MATLAB software. It should be noted that, with both models, 70% of data were randomly selected for training, 15% for validation, and 15% for testing the model.

Modeling the Data with the Decision Tree

For modeling the data using the decision tree method, the CART method was used. Figure 1 depicts the tree derived from this model. Table 2 shows the confusion matrix resulted from applying the decision tree to the research data.

Considering the confusion matrix, the values derived from evaluating the decision tree performance are as follows:

Accuracy = 70.3%

Sensitivity = 63.6%

Specificity = 73.1%

By definition, accuracy is the proportion of correctly classified data to the whole data. Since the data were slightly asymmetric, the number of samples labeled as class 0 was about 2.5 times more than the number of samples which belong to class 1. In this regard, two other scales were used to ensure optimum performance of the model. Sensitivity is the proportion of correctly classified data in class 1 to the whole data which were actually in class 1, and specificity represents the proportion of data labeled as class 0 to the whole data which were actually in class 0. Indeed, these two scales indicate how well the classification distinguishes between the two classes (18).
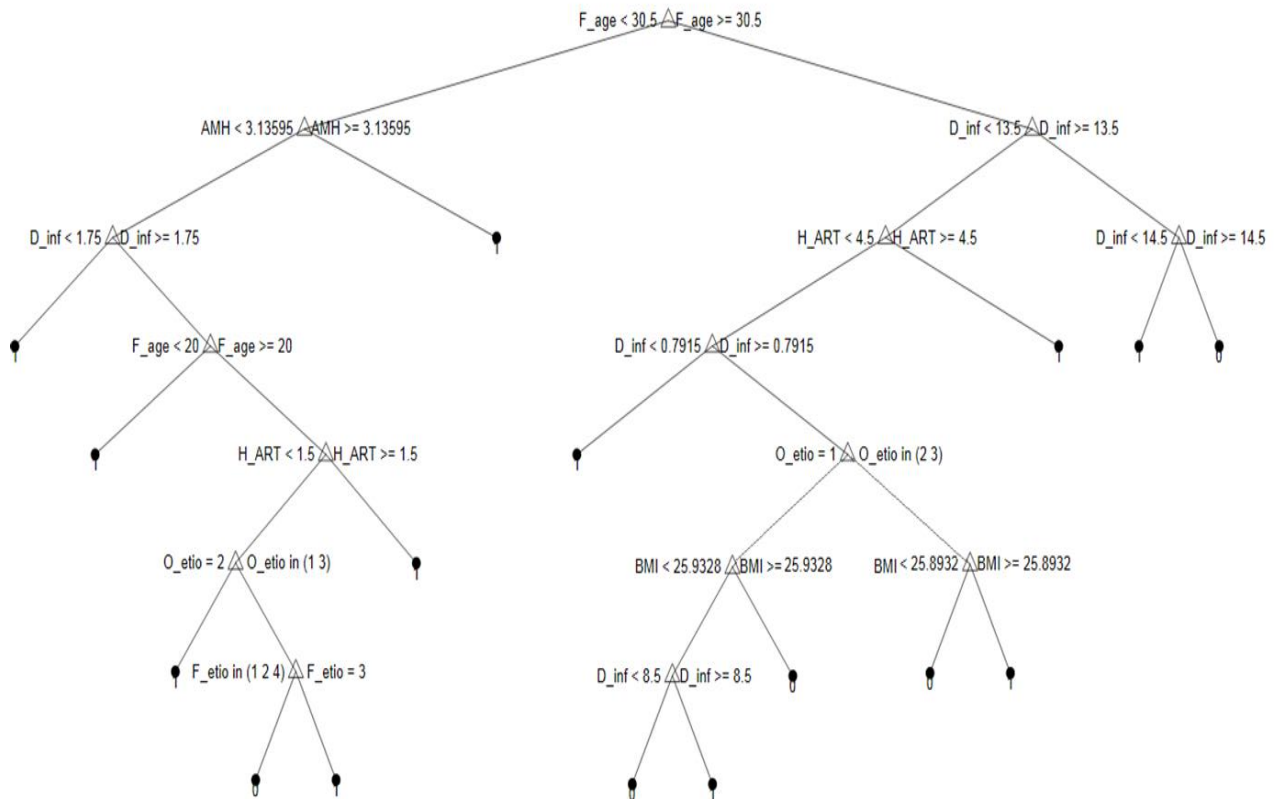

Figure 1. The tree obtained from the implementation of the model

Table 2. The confusion matrix resulted from applying the decision tree method

Simulated class \ True class	1	0
1	7	7
0	4	19

Modeling the Data with the Support Vector Machine Method

Table 3 shows the confusion matrix resulted from applying the support vector machine method on the data.

Considering the confusion matrix, the values derived from evaluating the support vector machine performance are as follows:

Accuracy = 75.7%

Sensitivity = 63.6%

Specificity = 80.8%

As indicated, the support vector machine performed better than the decision tree, and in general, it predicted the results of performing ICSI with higher accuracy.

Table 3. The confusion matrix resulted from applying the support vector machine method

Simulated class \ True class	1	0
1	7	5
0	4	21

Discussion

Studies carried out regarding infertility treatment have shown that there is still a long way to treat this disease satisfactorily. Spending so much time and money in the field of infertility treatment proves the necessity of designing a model which could be beneficial in predicting the outcomes of the treatment of this disease; a model that could predict the result of the treatment with acceptable accuracy, and accelerate the treatment process which could be cost-saving. Such a model helps doctors as well as patients to make the right decision regarding using or not using a treatment method.

In this paper, a model was proposed using the decision tree and support vector machine methods to predict the clinical pregnancy using ICSI. The input variables of the model were parameters like couple's medical records, hormonal tests, the cause of infertility, and the like. It is worth noting that, out of the nine input variables, ART and Anti-Mulirian Hormone have not been examined in previous studies; however, they were included in the current research as suggested by the supervising physician. One of the innovations of this study was that the input variables of the model were only preoperative, while, in the previous studies, having information about some stages of the surgery, such as quality of the egg, and the like, was required to anticipate the result of the surgery.

As mentioned, the ethical issues and confidential nature of patients' information is considered in this study. It was also noted that some studies were not successful in their data analyses because of failing to apply the effective parameters in the treatment method fully.

Conclusion

The results of the current study indicated that the support vector machine method performed better than the decision tree method and it predicted the output variables with higher accuracy. According to the results of data analysis, the proposed method could predict the output variable with an accuracy of 75.7%. Using this model can predicting the result of ICSI for an infertile couple and give it to them and

doctor; so as to decide whether to use or not to use this treatment method helping them. This will prevent some time and economic costs.

Given that having multiple pregnancies is one of the consequences of using the assisted reproductive technology, future studies are suggested to introduce a model which increases class labels to 4 (0, 1, 2, and 3) and predicts the number of embryos that remain until the clinical pregnancy. Developing such a system can help physicians to determine the number of embryos candidate for transfer; thus, it helps to control the consequence mentioned above. Furthermore, another research needs to be conducted in which the success of the surgery until giving birth to an infant is considered as an output variable.

Conflict of Interest

Authors declared no conflict of interests.

References

1. Rostami Dovom, M., Ramezani Tehrani, F., Abedini, M., Amirshekari, G., Hashemi, S., & Noroozzadeh, M. 2014. A population-based study on infertility and its influencing factors in four selected provinces in Iran (2008-2010). Iranian journal of reproductive medicine, 12(8), 561-566.

2. CORANI, G., MAGLI, C., GIUSTI, A., GIANAROLI, L. & GAMBARDELLA, L. 2013. A Bayesian network model for predicting pregnancy after in vitro fertilization. Computers in biology and medicine, 43, 1783-1792. [DOI:10.1016/j.compbiomed.2013.07.035] [PMID]

3. Otani, S., Iwai, T., Nakahata, S., Sakai, C., & Yamashita, M. 2009. Artificial fertilization by intracytoplasmic sperm injection in a teleost fish, the medaka (Oryzias latipes). Biology of reproduction, 80(1), 175-183. [DOI:10.1095/biolreprod.108.069880] [PMID]

4. GUH, R.-S., WU, T.-C. J. & WENG, S.-P. 2011. Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes. Expert Systems with Applications, 38, 4437-4449. [DOI:10.1016/j.eswa.2010.09.112]

5. SOHRABVAND, F., SHARIAT, M., FOTOOHI GHIAM, N. & HASHEMI, M. 2009. The relationship between number of transferred embryos and pregnancy rate in ART cycles. Tehran University Medical Journal, 67(2), 132-136.

6. ORVIETO, R., MELTCER, S., NAHUM, R., RABINSON, J., ANTEBY, E. Y. & ASHKENAZI, J. 2009. The influence of body mass index on in vitro fertilization outcome. International Journal of Gynecology & Obstetrics, 104, 53-55. [DOI:10.1016/j.ijgo.2008.08.012] [PMID]

7. PINBORG, A., GAARSLEV, C., HOUGAARD, C., NYBOE ANDERSEN, A., ANDERSEN, P., BOIVIN, J. & SCHMIDT, L. 2011. Influence of female bodyweight on IVF outcome: a longitudinal multicentre cohort study of 487 infertile couples. Reproductive BioMedicine Online, 23, 490-499. [DOI:10.1016/j.rbmo.2011.06.010] [PMID]

8. HAGHIGHI, Z., REZAEI, Z. & ES-HAGHI ASHTIANI, S. 2012. Effects of women's body mass index on in vitro fertilization success: a retrospective cohort study. Gynecological Endocrinology, 28, 536-539. [DOI:10.3109/09513590.2011.650657] [PMID]

9. Siristatidis, C. S., Chrelias, C., Pouliakis, A., Katsimanis, E., & Kassanos, D. 2010. Artificial neural networks in gynaecological diseases: Current and potential future applications. Medical Science Monitor, 16(10), RA231-RA236.

10. Siristatidis, C., Pouliakis, A., Chrelias, C., & Kassanos, D. 2011. Artificial intelligence in IVF: a need. System Biology in Reproductive Medicine, 57(4), 179-185. [DOI:10.3109/19396368.2011.558607] [PMID]

11. DORMAHAMMADI, S., ALIZADEH, S., ASGHARI, M. & SHAMI, M. 2014. Proposing a prediction model for diagnosing causes of infertility by data mining algorithms. Journal of Health Administration, 57(17), 46-57.

12. MILEWSKA, A.J., JANKOWSKA, D., CWALINA, U., CITKO, D., WIESAK, T., ACACIO, B. & MILEWSKI, R. 2015. Significance of discriminant analysis in prediction of pregnancy in IVF treatment. Studies in Logic, Grammar and Rhetoric, 43, 7-20. [DOI:10.1515/slgr-2015-0038]

13. MILEWSKI, R., JANKOWSKA, D., CWALINA, U., MILEWSKA, A.J., CITKO, D., WIESAK, T., MORGAN, A. & WOLCZYNSKI, S. 2016. Application of artificial neural networks and principal component analysis to predict results of infertility treatment using the IVF method. Studies in Logic, Grammar and Rhetoric, 47(1), 33-46. [DOI:10.1515/slgr-2016-0045]

14. MILEWSKI, R., KUCZYNSKA, A., STANKIEWICZ, B. & KUCZYNSKI, W. 2017. How much information about embryo implantation potential is included in morphokinetic data? A prediction model based on artificial neural networks and principal component analysis. Advances in Medical Sciences, 62(1), 202-206.

15. SEPEHRI, M.M., RAHNAMA, P., SHADPOUR, P. & TEIMOURPOUR, B. 2009. A data mining based model for selecting type of treatment for kidney stone patients. Tehran University Medical Journal, 67(6), 421-427.

16. HAN, J., KAMBER, M. & PEI, J. 2011. Data mining: concepts and techniques, Morgan kaufmann.

17. TAN, P.N., STEINBACH, M. & KUMAR, V. 2018. Introduction to data mining, Pearson Education India.

18. VIEIRA, S. M., MENDONCA, L. F., FARINHA, G. J. & SOUSA, J. 2013. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. Applied Soft Computing. [DOI:10.1016/j.asoc.2013.03.021]

How to Cite This Article:

Download citation:
BibTeX | RIS | EndNote | Medlars | ProCite | Reference Manager | RefWorks