Computational prediction of implantation outcome after embryo transfer

Behnaz Raef  
Tabriz University of Medical Sciences, Iran

Masoud Maleki  
Islamic Azad University, Iran

Reza Ferdousi  
Tabriz University of Medical Sciences, Iran

Abstract

The aim of this study is to develop a computational prediction model for implantation outcome after an embryo transfer cycle. In this study, information of 500 patients and 1360 transferred embryos, including cleavage and blastocyst stages and fresh or frozen embryos, from April 2016 to February 2018, were collected. The dataset containing 82 attributes and a target label (indicating positive and negative implantation outcomes) was constructed. Six dominant machine learning approaches were examined based on their performance to predict embryo transfer outcomes. Also, feature selection procedures were used to identify effective predictive factors and recruited to determine the optimum number of features based on classifiers performance. The results revealed that random forest was the best classifier (accuracy = 90.40% and area under the curve = 93.74%) with optimum features based on a 10-fold cross-validation test. According to the Support Vector Machine-Feature Selection algorithm, the ideal numbers of features are 78. Follicle stimulating hormone/human menopausal gonadotropin dosage for ovarian stimulation was the most important predictive factor across all examined embryo transfer features. The proposed machine learning-based prediction model could predict embryo transfer outcome and implantation of embryos with high accuracy, before the start of an embryo transfer cycle.

Corresponding author:
Reza Ferdousi, Department of Health Information Technology, School of Management and Medical Informatics, Tabriz University of Medical Sciences, Tabriz 5166/15731, Iran.
Email: ferdousir@tbzmed.ac.ir

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Keywords
assisted reproductive technology, embryo transfer, machine learning, prediction model, ranking algorithms

Introduction
Infertility is known as a disorder of the reproductive system diagnosed by the failure to conceive after 12 months or more of regular unprotected sexual intercourse. Based on the latest definition by the World Health Organization, “infertility is a disease which generates disability as an impairment of function.” Infertility is the most common global health complaint. More than 186 million couples worldwide suffer from infertility, and the majority of infertile couples are deprived from appropriate treatments in developing countries. There are many types of infertility treatment, including lifestyle changes (e.g. losing weight), medical treatments (e.g. use of drugs for ovulation induction), surgical treatments (e.g. laparoscopy), and assisted reproductive technologies (ARTs). ARTs are advanced technologies that human oocytes and sperm are handled and fertilized in in vitro conditions and embryos are transferred to the woman’s uterus for establishing a clinical pregnancy. An ART cycle, which is over an interval of approximately 2 weeks, involves several sequential steps that are complicated, time-consuming, costly, and hard to endure by infertile couples. The live birth rate for each complete ART cycle is 29.1 percent. However, 38 to 49 percent of couples will stay resultless, even after six treatment cycles. Therefore, opposite to the general belief, ART does not guarantee the success.

In vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI) are two popular ART procedures, which have nearly the same stages of process. The main difference between IVF and ICSI is the method of sperms and eggs fertilization. IVF is a conventional treatment including standard insemination of oocytes with sperms outside the body, while ICSI is an extension of IVF, which is performed by injection of a selected sperm cell into the oocyte cytoplasm. IVF is used for different causes of infertility, especially female factors, while ICSI is a proper treatment option in severe male factor infertility, such as azoospermia. An ART treatment cycle usually starts with controlled drug-induced ovarian stimulation to produce several mature eggs. Gonadotropins, such as human menopausal gonadotropin (HMG) and follicle-stimulating hormone (FSH), are used for stimulation in different protocols and prescriptions. Human chorionic gonadotropin (HCG) is also used as an ovulation trigger to stimulate final oocyte maturation before oocyte retrieval. With oocyte production, the cycle progresses to the oocyte retrieval phase for fertilization with sperm in the laboratory. After the eggs are fertilized, the resultant embryos are cultured, and then selected embryos at the cleavage or blastocyst stage are transferred into the woman’s uterus. Embryo transfer (ET) is the most critical stage in ART, which is composed of many variables, strategies, and techniques. All aspects, mentioned above, are important for overall ART success.

Reliable and accurate prediction of ART outcome is considered as an unsolved issue in the literature. Considering financial burden, physical and emotional risks, multiple pregnancies, complex process of treatment, and low rate of success, it is essential that infertile couples are well informed about their treatment with ART. On the other hand, there is a weak concordance between clinicians on treatment decisions and pregnancy probability estimation. To overcome these problems (i.e. predicting the probability of pregnancy), utilizing computational prediction models is an optimized solution. Clinical prediction models estimate the treatment results with a strong likelihood and also allow the treatment process to be adapted by using a variety of related parameters such as patient parameters and other effective ART cycle-specific variables.

To develop more accurate prediction models and algorithms with high-performance capacity, advanced computational approaches and data mining methods could be employed. The focus of
machine learning and data mining techniques is on developing computerized modeling and efficient predictive algorithms by detecting the hidden patterns in data to discover knowledge with high predictive accuracy. However, traditional statistical methods have focused on assessment and proportion of predefined models, which are not well fitted to face with mentioned complicated challenges.17,19,20

There are many types of machine learning approaches, such as Bayes nets, support vector machines (SVMs), decision trees (DTs), and so on.21 In machine learning methods, most of the input data are used for training algorithm(s). Indeed, the purpose of machine learning is to design and develop prediction models that enable the computer to solve a specific problem by learning from past data or experience.22 A practical application of machine learning in medicine is clinical decision support systems that allow the users, including related health professionals and patients, to indicate a suitable therapeutic plan.23,24

Machine learning methods are powerful computational tools to perform analysis in ART data and predict treatment outcomes. However, the literature on this field is limited.25 The previous studies rarely examined two aspects of embryological and clinical data together with and just considering a small number of effective variables.19,26 Moreover, the lack of categorization and ordering of the features in the previous studies led to difficult interpretation and comparison of the results.27

In a recent publication,28 we introduced 20 prediction models on ART. The prediction target of all these models is pregnancy (i.e. clinical pregnancy and ongoing pregnancy). However, prediction of ET outcome as a critical step is a major gap in the literature.

This study aims to construct a prediction model for ET outcome, using a comprehensive, varied features set and different machine learning algorithms. Several steps were performed to develop the ET predictor. The first step of this study was to identify predictive factors on ET, which is the combination of demographics, clinical, embryological, and ART cycle parameters related to infertile couples. The second step was to construct a prediction model by applying machine learning algorithms to build a computational model for predicting ET outcomes. In the last step of this study, we performed a comparative analysis of machine learning algorithms to determine which model(s) predicted ET outcomes with better performance in terms of their accuracy, sensitivity, specificity, and so on. Figure 1 represents the overall process of the study.

Materials and methods

Data acquisition and preparation

The data were obtained from 500 patients undergoing IVF and ICSI treatment at the East Azerbaijan ACECR ART center (Tabriz, Iran) from April 2016 to February 2018. In the obtained data of the 500 embryo transfer cases, 251 samples were recorded positive β-HCG and 249 samples were recorded negative β-HCG. In this study, we excluded other infertility treatment procedures at this clinic that does not include embryo transfer, such as intrauterine insemination. The complementary information about the recruited ET data set is given in Figure 2. This study and the data collection phase were approved by the Ethics Committee of Tabriz University of Medical Sciences (IR. TBZMED.REC.1397.345).

According to Figure 2, except 22 single embryo transfer (SET) cycles, more than one embryo were transferred to each patient through different cycles. Therefore, the total number of embryos transferred in the dataset was 1360.

To prepare the data set, first, all the data about ET stored in paper-based medical records were collected in an electronic format and then preprocessed. Data preprocessing, one of the critical
steps in machine learning was executed by handling missing values, outlier data, and application of normalization methods. The missing values of numerical features are replaced with median and categorical attributes filled by mode of their corresponding feature. To achieve an actual
and better outcome, records of patients with embryo donation and surrogacy, which could be resulting in misclassification and noise in prediction, were eliminated. Embryo donation and surrogacy are two phenomena of modern ART that introduce legal, ethical, and biological issues (e.g. genetic disparities of multiparents).34,35

**Attribute extraction and selection**

In this study, to predict implantation potential, all required variables were extracted from clinical-related guidelines, papers, and infertility specialists. At this point, after searching and obtaining relevant features, we developed a checklist and conducted a focus group with experts, including 14 obstetricians and gynecologists, two embryologists, two medical geneticists, and one social medicine physician. During this process, we collected expert’s opinions and attitudes about initial features set by a checklist based on Likert-type scale. Moreover, they could comment at the end of the checklist in response to an open question about other useful variables that are not in the variable list. Based on the expert’s opinions, we added some variables (i.e. anemia, thyroid disease, prolactin hormone disorders, amenorrhea, dysmenorrhea, period status, hirsutism, galactorrhea). After the survey and reevaluation of each element, the final feature set was selected as potential predictors for ET and implantation (Figure 1(a)).

Ultimate attributes consist of two main groups: (1) patient-related features (e.g. demographics, diagnostic, and clinical characteristics) and (2) ART cycle features (e.g. oocyte stimulation/morphology data and embryological data). Also, patient-related data were divided into female and male subcategories. The value of β-HCG was considered as a target variable (1 for positive β-HCG and 0 for negative β-HCG).

There were 59 and 23 attributes in groups 1 and 2, respectively. All of the recruited features have the potential to affect the performance of an algorithm. The features and their attribute types are summarized in Table 1.

For the prediction of embryo implantation, we examined six common ML algorithms on all groups of attributes. Furthermore, to identify effective features and their particular values that affect the outcome of embryo transfer, we applied feature selection (FS) and ranking algorithms.
Table 1. Description of extracted features in ET dataset.

| Attribute name                                     | Attribute type                  |
|---------------------------------------------------|--------------------------------|
| **Clinical data (patient-related data)**          |                                |
| Age of female                                     | Numeric                         |
| Age of male                                       | Numeric                         |
| BMI (body mass index)                             | Numeric                         |
| Family relation of couples                        | Categorical (yes, no)           |
| Family relation in parents of couples             | Categorical (yes, no)           |
| Smoking                                           | Categorical (yes, no)           |
| Type of infertility                               | Categorical (primary, secondary)|
| Infertility duration                              | Numeric                         |
| Contraception duration                            | Categorical (yes, no)           |
| Infertility in family                             | Categorical (yes, no)           |
| G (gravida/gravidity)                             | Numeric                         |
| P (para/parity)                                   | Numeric                         |
| Ab (abortion)                                     | Numeric                         |
| EP (ectopic pregnancy)                            | Numeric                         |
| L (living children)                               | Numeric                         |
| D (dead children)                                 | Numeric                         |
| Comorbidity diseases                              | Categorical (yes, no)           |
| Anemia                                            | Categorical (yes, no)           |
| Thyroid disease                                   | Categorical (hyper/hypo)        |
| Prolactin hormone disorders                       | Categorical (hyper/hypo)        |
| Drug usage                                        | Categorical (yes, no)           |
| **Female pathology data**                         |                                |
| Amenorrhea (absence of menstruation)             | Categorical (yes, no)           |
| Dysmenorrhea (painful periods)                    | Categorical (yes, no)           |
| Period status                                     | Categorical (yes, no)           |
| Hirsutism (excessive body hair in women)          | Categorical (yes, no)           |
| Galactorrhea (abnormal milky breast discharge)    | Categorical (yes, no)           |
| Gynecological surgery                             | Categorical (yes, no)           |
| Oocyte donation                                   | Categorical (yes, no)           |
| AFC (antral follicle count)                        | Categorical (normal, abnormal)  |
| Endometrium (tissue lining of the uterus) thickness| Numeric                         |
| Three-line (regular/normal) endometrium           | Categorical (yes, no)           |
| Uterus depth                                      | Numeric                         |
| Size of follicles                                 | Numeric                         |
| Tubal factor                                      | Categorical (yes, no)           |
| Pelvic factor                                     | Categorical (yes, no)           |
| Cervical factor                                   | Categorical (yes, no)           |
| Ovulatory factor                                  | Categorical (yes, no)           |
| PCOS (polycystic ovary syndrome)                  | Categorical (yes, no)           |
| Uterine factor                                    | Categorical (yes, no)           |
| Endometriosis (abnormal growth of endometrium in  | Categorical (yes, no)           |
| outside of the uterus cavity)                     |                                |
| Endometrial factor                                | Categorical (yes, no)           |
| Vaginitis                                         | Categorical (yes, no)           |
| RIF (repeated implantation failure)               | Categorical (yes, no)           |
| RPL (recurrent pregnancy loss)                    | Categorical (yes, no)           |

(Continued)
| Attribute name                                                                 | Attribute type                                                      |
|-------------------------------------------------------------------------------|---------------------------------------------------------------------|
| Thrombophilia disorders                                                       | Categorical (yes, no)                                               |
| Immunologic disorders                                                         | Categorical (yes, no)                                               |
| Male pathology data                                                           |                                                                     |
| Male factor                                                                   | Categorical (yes, no)                                               |
| Male genital surgery                                                           | Categorical (yes, no)                                               |
| Varicocele (abnormal enlargement of the testicular veins)                      | Categorical (yes, no)                                               |
| TESE (testicular sperm extraction)                                             | Categorical (yes, no)                                               |
| PESE (percutaneous epididymal sperm extraction)                                | Categorical (yes, no)                                               |
| Fresh/freeze sperm                                                            | Categorical (yes, no)                                               |
| Semen analysis data                                                            |                                                                     |
| Sperm count                                                                    | Numeric                                                             |
| Normal morph                                                                   | Numeric                                                             |
| Immotile                                                                       | Numeric                                                             |
| Lab tests                                                                      |                                                                     |
| FSH (follicle-stimulating hormone)                                            | Numeric                                                             |
| LH (luteinizing hormone)                                                       | Numeric                                                             |
| Estradiol                                                                      | Numeric                                                             |
| vitD3 Levels                                                                   | Categorical (deficiency, insufficiency, sufficiency)                 |
| Oocyte stimulation and morphology                                              |                                                                     |
| FSH/HMG (human menopausal gonadotropin) dosage                                 | Numeric                                                             |
| GnRH (gonadotropin-releasing hormone) antagonists Dosage                       | Numeric                                                             |
| GnRH agonists dosage                                                           | Numeric                                                             |
| Duration of stimulation (days)                                                 | Numeric                                                             |
| Estradiol dosage                                                               | Numeric                                                             |
| No. estradiol days                                                             | Numeric                                                             |
| Number of retrieved oocytes                                                     | Numeric                                                             |
| Number of MII (metaphase II) quality oocytes                                   | Numeric                                                             |
| Number of MI (metaphase I) quality oocytes                                     | Numeric                                                             |
| Number of GV (germinal vesicle) quality oocytes                                | Numeric                                                             |
| Number of degenerated quality oocytes                                          | Numeric                                                             |
| Quality of injected MII oocytes                                                 | Categorical (normal, abnormal)                                      |
| Embryological data                                                             |                                                                     |
| Number of 2PN (pronuclear)                                                     | Numeric                                                             |
| Number of developed embryos                                                    | Categorical (A, B, C, D)                                            |
| Quality of developed embryos                                                   | Categorical (normal, abnormal)                                      |
| Quality of vitelline space                                                     | Categorical (normal, abnormal)                                      |
| ET (embryo transfer) strategies                                                | Categorical (fresh, freeze)                                         |
| ET day                                                                         | Numeric                                                             |
| Number of transferred embryos                                                  | Numeric                                                             |
| Number of blastomeres                                                          | Numeric                                                             |
| Quality and stages of transferred embryos                                      | Categorical (A, B, C, D)                                            |
| Experience of ET                                                               | Categorical (yes, no)                                               |
| PRP (platelet-rich plasma)                                                     | Numeric                                                             |
| ID (identification)                                                            | Meta                                                                |
| β-HCG (human chorionic gonadotropin)                                           | Target (positive, negative)                                         |
FS algorithms were used to identify and mine attributes that have a significant prognostic effect on implantation outcome. Hence, the relative weights for each feature were extracted. Classification algorithms were separately tested with the different number of weighted attributes to determine a set of features with superior performance. Therefore, the selected subset of features was used in the rest of the experiments.

**Implementing predictors with machine learning algorithms**

In this step, six ML algorithms were fed with preprocessed data to determine their performances in ET outcome prediction. Comparative analysis of diverse classifiers enabled us to determine the best fitting models for the employed data set. To predict ET outcome, six of the most well-known prediction algorithms, including SVMs, neural network (NN), k-nearest neighbors (kNN), naive bayes (NB), random forest (RF), and DT, were used to develop our predictor. The hypotheses underlying each of these algorithms were the minimization of empirical risk and reducing the errors in the training set. These classifiers were chosen and tested with Orange data mining software. Figure 1(b) shows the overall stages of the proposed machine learning-based ET outcome prediction model.

**Prediction assessment**

A standard assessment method was essential to evaluate the performance of each algorithm. In other words, the prediction process needs two types of data: training and testing data. In this study, 80 percent of the data set was used as a training set, and the remaining 20 percent was used as a test set. To do this, we used 10-fold cross-validation to assess the robustness of the approaches. The ET dataset was randomly divided into 10 equal-sized subsets, and the cross-validation process was repeated 10 times. Each time, one of the 10 subsets is used as the validation set for testing the model and the remaining nine subsets are put together to form a training data set. Finally, 10 results of experiments were averaged to produce a single estimation for each algorithm.

The performance of the algorithms was evaluated in terms of common standard machine learning evaluation parameters. These parameters were computed based on the values of true negatives (TN), true positives (TP), false positives (FP), and false negatives (FN) as detailed below.

- **Accuracy (ACC):** percentage of positive and negative β-HCG that was correctly predicted

  \[
  \text{ACC} = \frac{TP + TN}{(TP + FP + TN + FN)} \times 100
  \]

- **Sensitivity (SN):** percentage of positive β-HCG that was predicted correctly

  \[
  \text{SN} = \frac{TP}{(TP + FN)} \times 100
  \]

- **Specificity (SP):** percentage of negative β-HCG that was correctly predicted

  \[
  \text{SP} = \frac{TN}{(TN + FP)} \times 100
  \]
Matthew’s correlation coefficient (MCC): this value ranges from −1 for worst prediction to +1 for accurate prediction; 0 indicates random prediction

\[
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TN + FN) \times (TP + FN) \times (TN + FP)}} \times 100
\]

- Precision or positive predictive value (PPV)

\[
\text{PPV} = \frac{TP}{(TP + FN)} \times 100
\]

- Negative predictive value (NPV)

\[
\text{NPV} = \frac{TN}{(TN + FN)} \times 100
\]

- F-measure: this parameter is a combined evaluation of precision and recall

\[
\text{F-score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100
\]

- The area under the curve (AUC): this parameter is a logical evaluation for model performance. It is the value ranges from 0 to 1, where 1 represents the best performance, and 0 is the worst performance. AUC = 0.5 when random ranking is used.

Results

Attribute selection and predictive features

As mentioned before, for accurate prediction of ET outcome, the feature set with the optimum quantity and quality is essential; thus, more precise attribute selection methods are undoubtedly required. Among the different ranking algorithms, the SVM-FS resulted in better performance, so this FS algorithm was used to select and rank the optimal number of attributes. Based on SVM-FS, 78 features were chosen as the optimal feature set.

According to estimated feature weights, the FSH/HMG dosage was identified as the most effective predictor variable, and also the contraception duration and the number of germinal vesicle (GV) quality oocytes are other key features in the success of an ET cycle. On the other hand, the quality of injected metaphase II (MII) oocyte, sperm count, and male factor features have less predictive value on the ET outcome. The ranking and relative weights of these attributes that were calculated based on the SVM-FS algorithm are given in Table 2.

Classifier selection and predictive modeling

Six machine learning algorithms were employed to develop a model to predict ET outcome. Features were ranked by SVM-FS and after examining the performance of the algorithms with
Table 2. Scored variables related to the embryo transfer cycle, with SVM-FS.

| Feature name                                      | Weight | Feature name                  | Weight |
|---------------------------------------------------|--------|-------------------------------|--------|
| FSH/HMG dosage                                    | 8.685  | Gynecological surgery         | 0.077  |
| Contraception duration                            | 3.937  | Infertility duration          | 0.076  |
| Number of GV quality oocytes                      | 1.458  | vitD3 deficiency               | 0.073  |
| Uterus depth                                      | 1.048  | LH                            | 0.073  |
| Number of MII quality oocytes                     | 1.008  | G                             | 0.071  |
| RPL (recurrent pregnancy loss)                    | 0.935  | Smoking (male)                | 0.063  |
| Varicocele                                        | 0.875  | Three line endometrium        | 0.060  |
| RIF (repeated implantation failure)               | 0.836  | Thrombophilic disorders       | 0.056  |
| Fresh/freezing ET                                 | 0.741  | Hyper/hypoprolactinemia       | 0.047  |
| Number of 2PN                                     | 0.626  | Experience of ET              | 0.044  |
| GnRH agonists dosage                              | 0.616  | AFC                           | 0.042  |
| Estradiol level                                   | 0.579  | Age of male                   | 0.038  |
| FSH                                               | 0.526  | Patient diseases (comorbidity diseases) | 0.035 |
| Amenorrhea                                        | 0.498  | Embryo transfer day           | 0.033  |
| Number of blastomeres                              | 0.476  | Family relation of couples    | 0.029  |
| Hirsutism                                         | 0.416  | Ep                            | 0.025  |
| Infertility in family                              | 0.382  | Number of developed embryos   | 0.025  |
| Male genital surgery                               | 0.352  | Quality and stages of transferred embryos | 0.022 |
| Galactorrhea                                       | 0.331  | PESE                          | 0.020  |
| Endometrium thickness                              | 0.328  | Quality of vitelline space    | 0.020  |
| COC No.                                           | 0.322  | Ovulatory factor              | 0.019  |
| BMI                                               | 0.298  | Fresh/freeze sperm            | 0.017  |
| Size of follicles                                  | 0.267  | L                             | 0.013  |
| Thyroid                                           | 0.254  | Anemia                        | 0.013  |
| Number of transferred embryos                      | 0.244  | Tubal factor                  | 0.010  |
| PRP                                               | 0.239  | Duration of stimulation (days)| 0.005  |
| Uterine factor                                    | 0.210  | Type of infertility: primary/secondary | 0.004 |
| Age of female                                     | 0.173  | P                             | 0.003  |
| Estradiol/equine dosage                           | 0.161  | TESE                          | 0.003  |
| Cervical factor                                    | 0.154  | GnRH antagonists dosage       | 0.003  |
| Quality (grades) of developed embryos             | 0.149  | No. of days of estradiol usage| 0.003  |
| PCOS                                              | 0.136  | Family relation in parents of couples | 0.002 |
| Sperm morphology: Normal morph (%)                | 0.136  | Drug                          | 0.002  |
| Immotile sperm (%)                                 | 0.136  | Number of degenerated quality oocytes | 0.002 |
| Vaginitis                                         | 0.135  | Period status                 | 0.001  |
| Ab                                                | 0.109  | D                             | 0.001  |
| Number of oocytes of MI quality                    | 0.109  | Endometrial factor            | 0.001  |
| Dysmenorrhea                                       | 0.108  | Oocyte donation               | 0.001  |
| Pelvic factor                                      | 0.106  | Male factor                   | 0.001  |
| Endometriosis                                     | 0.095  | Sperm count                   | 0.000  |
| Immunologic disorders                              | 0.093  | Quality of injected MII oocyte| 0.000  |

BMI: body mass index; AFC: antral follicle count; PCOS: polycystic ovary syndrome; RIF: repeated implantation failure; RPL: recurrent pregnancy loss; TESE: testicular sperm extraction; PESE: percutaneous epididymal sperm extraction; FSH: follicle-stimulating hormone; LH: luteinizing hormone; HMG: human menopausal gonadotropin; MII: metaphase II; GV: germinal vesicle; ET: embryo transfer; PRP: platelet-rich plasma.
different number of features; finally, 78 features were selected. The performance of each algorithm, without and with FS is summarized in Table 3.

As highlighted in Table 3, the best classification performance based on the highest CA and AUC belonged to the NN and RF classifiers. Therefore, among the six classification algorithms, the NN and RF were considered as the best algorithms for ET data. Selected features after FS (78 features) in comparison with all features (82 features without FS) showed better performance (Table 3).

Figure 3 shows the AUC plots of different algorithms. The performances of the RF and NN are slightly better both before and after FS.

Discussion

The present study showed that machine learning methods could help to predict the implantation outcomes of embryo transfers by determining the essential factors on the ART treatment procedure. Employing more relevant attributes is crucial for building a functional and straightforward prediction model with high performance. Therefore, we applied an adequate ET data set that included comprehensive and detailed features of patient demographics, embryo parameters, and cycle characteristics with a sufficient number of records to train a model by powerful machine learning prediction methods. In the reviewed literature, the maximum number of recruited features in the ART outcome prediction models was 64; however, the minimum number of features was four in Wald et al.

To the best of our knowledge, few previous studies related to ART outcome predictions concentrated on only a small number of attributes and limited aspects, which are unlikely to represent all effective factors on embryo implantation. Only five papers of 20 ART outcome prediction models used an almost comprehensive features set. However, we increased the number of features, such as semen analysis parameters, differently with earlier studies. In this study, to achieve high prediction accuracy, pivotal factors in ART were considered as different feature groups and analyzed in detail (Table 1).

Remarkably, covering a more extensive range of embryo morphological data including both cleavage-stage (i.e. day 2 and day 3) and blastocyst-stage embryos (i.e. day 5 or day 6 or more) is another important distinction of this study. However, the development of a model with combined embryo patterns of cleavage and blastocyst stages was mentioned as a limitation, and further investigation is suggested in earlier studies. Also, most of them have focused on cleavage stages-related parameters, while blastocyst transfers provide a higher implantation rate compared with the cleavage stages.

There is a growing interest in embryologists to transfer multiple embryos to increase implantation chance and obtain high pregnancy rates, which resulted in high rates of multiple pregnancies and births. The various pregnancies have been documented as a significant public health issue that leads to many maternal and neonatal complications and risks. Therefore, the SET strategy has been recognized worldwide as the only practical solution to overcome this problem and avoid multiple pregnancies in ART cycles, and many countries establish rules for encouraging or mandating increased use of SET. Predictive models with decision support capabilities that are based on embryo assessment parameters may facilitate better selection of embryos with the highest implantation potential and facilitate the utilization of a SET policy.

It is crucial to determine which features have a potential role in the prediction of ART treatment outcome. In many of previous studies, the age of woman is the most important attribute in prediction of ART outcomes. According to the results obtained from the present study,
Table 3. Performance of six algorithms without and with feature selection (82 vs. 78 features).

| Number of features | Classifier | CA  | SN  | SP  | AUC | IS  | F1  | PPV | Recall | Brier | MCC  |
|--------------------|------------|-----|-----|-----|-----|-----|-----|-----|--------|-------|------|
| 82                 | NB         | 83.60 | 81.93 | 85.26 | 87.97 | 63.98 | 83.27 | 84.65 | 81.93 | 29.72 | 67.23 |
|                    | SVM        | 80.60 | 76.31 | 84.86 | 87.10 | 43.09 | 79.66 | 83.33 | 76.31 | 29.11 | 61.40 |
|                    | NN         | 89.60 | 85.94 | 93.23 | 93.63 | 73.05 | 89.17 | 92.64 | 85.94 | 17.42 | 79.40 |
|                    | RF         | 89.60 | 90.04 | 89.16 | 93.68 | 61.93 | 89.68 | 89.33 | 90.04 | 19.16 | 79.20 |
|                    | KNN        | 84.20 | 74.30 | 94.02 | 92.66 | 65.42 | 82.41 | 92.50 | 74.30 | 23.80 | 69.73 |
|                    | DT         | 87.80 | 83.13 | 92.43 | 89.05 | 73.85 | 87.16 | 91.59 | 83.13 | 23.16 | 75.91 |
| 78                 | NB         | 83    | 80.72 | 85.26 | 88.09 | 63.10 | 82.55 | 84.45 | 80.72 | 30.50 | 66.06 |
|                    | SVM        | 80.20 | 77.11 | 83.27 | 86.80 | 42.87 | 79.50 | 82.05 | 77.11 | 29.30 | 60.50 |
|                    | NN         | 90.40 | 87.95 | 92.83 | 93.47 | 73.31 | 90.12 | 92.41 | 87.95 | 17.04 | 80.89 |
|                    | RF         | 90.40 | 90.36 | 90.44 | 93.74 | 62.35 | 90.36 | 90.36 | 17.84 | 80.80 |
|                    | KNN        | 83.80 | 74.70 | 92.83 | 92.14 | 64.62 | 82.12 | 91.18 | 74.70 | 24.43 | 68.70 |
|                    | DT         | 88.60 | 83.94 | 93.23 | 89.14 | 76.08 | 88    | 92.48 | 83.94 | 21.70 | 77.52 |

NB, naive bayes; SVM, support vector machine; NN, neural network; RF, random forest; KNN, k-nearest neighbor; DT, decision tree; CA, classification accuracy; SN, sensitivity; SP, specificity; AUC, area under the curve; IS, information score; F1, F-measure; PPV, precision or positive predictive value; MCC, Matthew’s correlation coefficient.

Figure 3. Area under the curve (AUC) plots of different algorithms: (a) all 82 attributes and (b) the optimum number of features selected by SVM-FS.

FSH/HMG dosage was a high weighted feature. This feature is related to the controlled ovarian stimulation (COH) phase, which is an initial procedure in the ART cycle to induce the growth of follicles with gonadotropins. It is in accordance with an earlier study that remarked the importance of the total dose of gonadotropins with different COH treatment protocols on the adequate number of retrieved eggs in the success of IVF.47
Contraception duration was the other crucial prognostic factor determined by this study. In earlier studies, the various contraceptive methods have different health risks, and infertility is one substantial adverse effect on them. It is founded that long-term use of birth control methods, for example, intrauterine device, is associated with increased risk of fertility disability, by causing different types of infertility in women, such as tubal or ovulatory causes.48–50

The female oocytes play a crucial role in the developmental competence of embryo and later on ART results.51 The immature oocytes result in a significant reduction in IVF success. Therefore, the oocytes that are arrested at the GV stage are incompetent oocytes for fertilization with sperms and embryo development.52 On the other hand, usually all MII oocytes, as mature oocytes and ready for fertilization, are collected and inseminated.53 Hence, progressing the maturation cycle to MII phase and considering their number per ART cycle play an essential role in the chance of establishing pregnancy.54 These are following our feature ranking result that introduced the number of GV and MII quality oocytes as potential factors in the prediction of ET outcome. Results of feature ranking also show that uterus depth is another effective feature in ET outcome. This finding has been confirmed by previous studies that reported the impact of the uterine cavity measure for suitable ET and subsequent pregnancy rate in ART cycles.55

A straight comparison of the presented results in this study with those in the literature is not possible because of the diversity of research purposes, input data, applied analytical software, algorithms, and strategies for training/testing that play a vital role in the performance of model and selection of highly effected predictive features.

In this study, we used six common machine learning algorithms to develop a prediction model for ET outcomes. The performance of each algorithm was determined by evaluating how correctly they could predict whether embryos were implanted or not implanted, and the gold standards of evaluation metrics were used. The area under the receiver-operating characteristic (ROC) curve is accepted as a reliable and popular performance measure for assessing the quality of classification algorithms in machine learning approaches.18 The high value of AUC in this study (Table 3) shows the reliability of the presented approach for ET prediction.

The results showed that the prediction performance improved by applying FS and classification threshold optimization. Among the implemented algorithms, NN and RF showed superior performance to NB, SVM, kNN, and DT models, particularly by using a reduced and ranked feature set and optimized threshold of sampling (10-fold cross-validation). The NN algorithm has been used in two studies46,56 as single technique, and in other studies, it has been used along with other algorithms.6,25,37,39,57–60 In accordance with our results, the NN algorithm in comparison with other algorithms has been selected as a more suitable method in ART outcome prediction.37,39,59 Another superior algorithm in our study was RF. In support of this result, in three studies,6,40,60 among the five studies using the RF algorithm along with others, the RF was identified as a better method. Also, in previous studies18,39 the RF algorithm has achieved performance close to that of the superior algorithm. More studies in this domain have applied approaches based on classical statistical methods, such as logistic regression. A few rare studies in the literature were performed based on machine learning techniques that were different from current research in terms of aim and target variable, the number of feature sets, recruited algorithms, and results.

In this study, we faced several limitations. One of the restrictions was generalization of our the proposed model, since the input data in this work come from a single source, and thus may yield that this model works in a given clinic but cannot necessarily be transferred to other clinics without adaptation of the algorithm parameters to clinic-specific characteristics (e.g. culture conditions, fertilization method, media). Another fundamental limitation was the absence of any electronic documentation at the investigation center that resulted in many problems and significant time consumption on data gathering and data entry. Illegibility of paper-based records, incomplete patients
records, and missing values affected the performance of classifiers and FS algorithms. Due to these limitations, we had to eliminate some important variables such as culture medium because majority of values were missed. Due to the lack of electronic dataset and unavailability of public registries, the data sharing in this domain is one of the major challenges.25

Using implantation and primary β-HCG test (after embryo transfer) versus live birth as an endpoint for the model development provides the opportunity of investigating different variables in the ART cycle. However, a positive β-HCG does not grantee a live birth, which is not focused in this study. Also in this study, β-HCG values collected from patient records were not homogeneous (different laboratories and immunoassay methods led to a different level of β-HCG),61 which may influence the accuracy of the presented model. The issues mentioned above also had been documented in earlier studies.6,25,44

This study shows that the combination of different subsets of ET attributes and efficient machine learning algorithms can significantly improve and boost the predictability of the embryo implantation outcome. Hence, the results of prediction by the model could help make more accurate decisions in embryo selection by embryologists and minimize the current challenges in ART treatments. The proposed model can reduce costs of ART treatments by preventing repeated ART cycles. High expenditures of ART cycles is one of the major barriers that have significant economic effects on communities.62,63 Unlike high expenses, IVF/ICSI are not covered by health insurance.64

Conclusion
The application of computational approaches in the prediction of human embryo implantation can increase pregnancy rate after ART treatments. The proposed machine learning-based model can provide a clinical decision support tool to clinicians and infertile couples to consider the chances of success before the treatment procedure. Since this model integrates the experiences of all experts and the history of treatments into a single computational tool, learns from the past cases, and analyzes several embryos and patient records, it can make predictions in minimum time with less subjectivity, human bias, and higher precision. Also, using such an intelligent model is expected has a promising benefit in the selection of the best embryo with the highest implantation potential to transfer in IVF/ICSI treatment and may be used as an educational assistant tool for embryologists.

Our findings support the possibility and benefits of these applications, and the results of the actual use of this tool in clinical practice by a prospective trial are highly valuable. As future work, collecting similar datasets from various infertility clinics and applying the proposed predictive model for covering a wider range of ET data distributions toward external validation and impact analysis of the model is recommended. A further extension of this study could focus on the ultimate goal of treatment (healthy birth after positive β-HCG). Rich datasets with the variability of intervening parameters in this domain, such as embryological factors (i.e. genetic screening, time-lapse monitoring, and morphokinetic parameters of the embryo), and further examinations on the female/male partner (i.e. metabolomics and extra hormonal tests) could improve the accuracy of this model.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was a part of an MSc thesis that approved and funded by Tabriz University of Medical Sciences.

ORCID iD
Reza Ferdousi https://orcid.org/0000-0001-5511-5452

References
1. Zegers-Hochschild F, Adamson GD, de Mouzon J, et al. The International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary on ART terminology, 2009. *Hum Reprod* 2009; 24: 2683–2687.
2. Zegers-Hochschild F, Adamson GD, Dyer S, et al. The international glossary on infertility and fertility care, 2017. *Fertil Steril* 2017; 108: 393–406.
3. Inhorn MC and Patrizio P. Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century. *Hum Reprod Update* 2015; 21(4): 411–426.
4. National Institute for Health and Clinical Excellence (NICE). *Fertility problems: assessment and treatment*. London: NICE, 2017.
5. Sunderam S, Kissin DM, Crawford SB, et al. Assisted reproductive technology surveillance-United States, 2011. *MMWR Surveill Summ* 2014; 63(10): 1–28.
6. Hafiz P, Nematollahi M, Boostani R, et al. Predicting implantation outcome of in vitro fertilization and intracytoplasmic sperm injection using data mining techniques. *Int J Fertil Steril* 2017; 11(3): 184–190.
7. McLernon D, Steyerberg E, te Velde E, et al. 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.
8. Dhillon RK, McLernon DJ, Smith PP, et al. Predicting the chance of live birth for women undergoing IVF: a novel pretreatment counselling tool. *Hum Reprod* 2016; 31(1): 84–92.
9. Gozlan I, Dor A, Farber B, et al. Comparing intracytoplasmic sperm injection and in vitro fertilization in patients with single oocyte retrieval. *Fertil Steril* 2007; 87(3): 515–518.
10. Ou YC, Lan KC, Huang FJ, et al. Comparison of in vitro fertilization versus intracytoplasmic sperm injection in extremely low oocyte retrieval cycles. *Fertil Steril* 2010; 93(1): 96–100.
11. Eftekhari M, Mohammadian F, Yousefnejad F, et al. Comparison of conventional IVF versus ICSI in non-male factor, normoresponder patients. *Iran J Reprod Med* 2012; 10(2): 131–136.
12. Van Steirteghem A, Devroey P and Liebaers I. Intracytoplasmic sperm injection. *Mol Cell Endocrinol* 2002; 186: 199–203.
13. National Collaborating Centre for Women’s Children’s Health (NCC-WCH). *Fertility: assessment and treatment for people with fertility problems*. 2nd ed. London: RCOG Press, 2013.
14. Practice Committee of the American Society for Reproductive Medicine (ASRM). Performing the embryo transfer: a guideline. *Fertil Steril* 2017; 107: 882–896.
15. Corani G, Magli C, Giusti A, et al. A Bayesian network model for predicting pregnancy after in vitro fertilization. *Comput Biol Med* 2013; 43(11): 1783–1792.
16. Van der Steeg JW, Steures P, Eijkemans MJ, et al. Do clinical prediction models improve concordance of treatment decisions in reproductive medicine. *BJOG* 2006; 113(7): 825–831.
17. Malinowski P, Milewski R, Ziniewicz P, et al. The use of data mining methods to Predict the Result of Infertility Treatment Using the IVF ET Method. *Stud Logic Gram Rhet* 2014; 39: 67–74.
18. Guvenir HA, Mısırlı G, Dilbaz S, et al. Estimating the chance of success in IVF treatment using a ranking algorithm. *Med Biol Eng Comput* 2015; 53(9): 911–920.
19. Chen F, De Neubourg D, Debrock S, et al. Selecting the embryo with the highest implantation potential using a data mining based prediction model. *Reprod Biol Endocrinol* 2016; 14: 10.
20. Shmueli G. To explain or to predict? *Stat Sci* 2010; 25: 289–310.
21. Leskovec J, Rajaraman A, Ullman JD, et al. *Mining of massive datasets*. Cambridge: Cambridge University Press, 2014.

22. Alpaydin E. *Introduction to machine learning*. 2nd ed. Cambridge, MA: The MIT Press, 2014.

23. Shouval R, Bondi O, Mishan H, et al. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone Marrow Transplant* 2013; 49: 332–337.

24. Rezaei-Hachesu P, Oliyae A, Saffaie N, et al. Comparison of coronary artery disease guidelines with extracted knowledge from data mining. *J Cardiovasc Thorac Res* 2017; 9(2): 95–101.

25. Uyar A, Bener A and Ciray HN. Predictive modeling of implantation outcome in an in vitro fertilization setting: an application of machine learning methods. *Med Dec Mak* 2014; 35: 714–725.

26. Holte J, Berglund L, Milton K, et al. Construction of an evidence-based integrated morphology cleavage embryo score for implantation potential of embryos scored and transferred on day 2 after oocyte retrieval. *Hum Reprod* 2007; 22(2): 548–557.

27. Van Loendersloot L, Repping S, Bossuyt PM, et al. Prediction models in in vitro fertilization; where are we? A mini review. *J Adv Res* 2014; 5(3): 295–301.

28. Raef B and Ferdousi R. A review of machine learning approaches in assisted reproductive technologies. *Acta Inform Medica* 2019; 27: 205–211.

29. Hu X. DB-HReduction: a data preprocessing algorithm for data mining applications. *Appl Math Lett* 2003; 16: 889–895.

30. Guh R-S, Wu T-CJ and Weng S-P. Integrating genetic algorithm and decision tree learning for assistance in predicting in vitro fertilization outcomes. *Exp Syst Appl* 2011; 38: 4437–4449.

31. Han J, Kamber M and Pei J. *Data mining concepts and techniques* (The Morgan Kaufmann Series in Data Management Systems) (3rd ed.). Burlington, MA: Morgan Kaufmann, 2011, pp. 83–124.

32. Zhang Z. Missing data imputation: focusing on single imputation. *Ann Transl Med* 2016; 4(1): 9.

33. Glas CAW. Missing data. In: Peterson P, Baker E and Mcgaw B (eds) *International encyclopedia of education* (3rd ed.). Oxford: Elsevier, 2010, pp. 283–288.

34. Aluas M. Ethical issues raised by multiparents. In: Hostiuc S (ed.) *Clinical ethics at the crossroads of genetic and reproductive technologies*. Oxford: Academic Press, 2018, pp. 81–97.

35. Reilly PR. Legal issues in genetic medicine. In: Rimoin D, Pyeritz R and Korff B (eds) *Emery and Rimoin’s principles and practice of medical genetics*. Oxford: Academic Press, 2013, pp. 1–15.

36. Jamali AA, Ferdousi R, Razzaghi S, et al. DrugMiner: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. *Drug Discov Today* 2016; 21(5): 718–724.

37. Wald M, Sparks A, Sandlow J, et al. Computational models for prediction of IVF/ICSI outcomes with surgically retrieved spermatozoa. *Reprod Biomed Online* 2005; 11(3): 325–331.

38. Kim I-C and Jung Y-G. Using Bayesian networks to analyze medical data. In: *International workshop on machine learning and data mining in pattern recognition*, Leipzig, 5–7 July 2003, pp. 317–327. Berlin: Springer.

39. Hassan MR, Al-Insiaf S, Hossain MI, et al. A machine learning approach for prediction of pregnancy outcome following IVF treatment. *Neur Comput Appl* 2020; 32: 2283–2297.

40. Blank C, Wildeboer RR, DeCroo I, et al. Prediction of implantation after blastocyst transfer in in vitro fertilization: a machine-learning perspective. *Fertil Steril* 2019; 111(2): 318–326.

41. Kirkegaard K, Agerholm IE and Ingerslev HJ. Time-lapse monitoring as a tool for clinical embryo assessment. *Hum Reprod* 2012; 27(5): 1277–1285.

42. Glujovsky D, Blake D, Bardach A, et al. Cleavage stage versus blastocyst stage embryo transfer in assisted reproductive technology. *Cochrane Database Syst Rev* 2012; 2012: CD002118.

43. Vaegter KK, Lakić TG, Olovsson M, et al. 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. *Fertil Steril* 2017; 107(3): 641.e2–648.e2.

44. Petersen BM, Boel M, Montag M, et al. Development of a generally applicable morphokinetic algorithm capable of predicting the implantation potential of embryos transferred on Day 3. *Hum Reprod* 2016; 31(10): 2231–2244.

45. Adashi EY, Barri PN, Berkowitz R, et al. Infertility therapy-associated multiple pregnancies (births): an ongoing epidemic. *Reprod Biomed Online* 2003; 7(5): 515–542.
46. Kaufmann SJ, Eastaugh JL, Snowden S, et al. The application of neural networks in predicting the outcome of in-vitro fertilization. *Hum Reprod* 1997; 12(7): 1454–1457.
47. Pandian Z, McTavish AR, Aucott L, et al. Interventions for “poor responders” to controlled ovarian hyper stimulation (COH) in in-vitro fertilisation (IVF). *Cochrane Database Syst Rev* 2010; 2010: CD004379.
48. Lee NC, Peterson HB, Chu SY, et al. Health effects of contraception. In: Parnell AM; National Research Council (US) Committee on Population (eds) *Contraceptive use and controlled fertility: health issues for women and children background papers*. Washington, DC: National Academies Press, 1989, pp. 48–95.
49. Doll H, Vessey M and Painter R. Return of fertility in nulliparous women after discontinuation of the intrauterine device: comparison with women discontinuing other methods of contraception. *BJOG* 2001; 108(3): 304–314.
50. Chasen-Taber L, Willett WC, Stampfer MJ, et al. Oral contraceptives and ovulatory causes of delayed fertility. *Am J Epidemiol* 1997; 146(3): 258–265.
51. Rienzi L, Balaban B, Ebner T, et al. The oocyte. *Hum Reprod* 2012; 27(Suppl. 1): i2–i21.
52. Chen ZQ, Ming TX and Nielsen HI. Maturation arrest of human oocytes at germinal vesicle stage. *J Hum Reprod Sci* 2010; 3(3): 153–157.
53. Ozgur K, Bulut H, Berkkanoglu M, et al. Oocyte maturation-index as measure of oocyte cohort quality; a retrospective analysis of 3135 ICSI cycles. *Mid East Fertil Soc J* 2015; 20: 37–42.
54. McAvey B, Zapantis A, Jindal SK, et al. How many eggs are needed to produce an assisted reproductive technology baby: is more always better? *Fertil Steril* 2011; 96(2): 332–335.
55. Madani T, Ashrafi M, Abadi AB, et al. Appropriate timing of uterine cavity length measurement positively affects assisted reproduction cycle outcome. *Reprod Biomed Online* 2009; 19(5): 734–736.
56. Durairaj M and Nandhakumar R. Data mining application on IVF data for the selection of influential parameters on fertility. *Int J Eng Adv Technol* 2013; 2: 262–2662.
57. Chen C-C, Hsu C-C, Cheng Y-C, et al. Knowledge discovery on in vitro fertilization clinical data using particle swarm optimization. In: Proceedings of the 9th international conference on bioinformatics and bioengineering, Taichung, Taiwan, 22–24 June 2009, pp. 278–283. New York: IEEE.
58. Nanni L, Luminia A and Manna C. A data mining approach for predicting the pregnancy rate in human assisted reproduction. In: Brahnam S and Jain LC (eds) *Advanced computational intelligence paradigms in healthcare* 5. Berlin: Springer, 2010, pp. 97–111.
59. Milewski R, Milewska AJ, Więsak T, et al. Comparison of artificial neural networks and logistic regression analysis in pregnancy prediction using the in vitro fertilization treatment. *Stud Logic Gram Rhet* 2013; 35: 39–48.
60. Mirroshandel SA, Ghasemian F and Monji-Azad S. Applying data mining techniques for increasing implantation rate by selecting best sperms for intra-cytoplasmic sperm injection treatment. *Comput Meth Prog Biomed* 2016; 137: 215–229.
61. Cao Z and Rej R. Are laboratories reporting serum quantitative hCG results correctly? *Clin Chem* 2008; 54(4): 761.
62. Teoh PJ and Maheshwari A. Low-cost in vitro fertilization: current insights. *Int J Womens Health* 2014; 6: 817–827.
63. Bouwmans CA, Lintsen BM, Eijkemans MJ, et al. A detailed cost analysis of in vitro fertilization and intracytoplasmic sperm injection treatment. *Fertil Steril* 2008; 89(2): 331–341.
64. Katz P, Showstack J, Smith JF, et al. Costs of infertility treatment: results from an 18-month prospective cohort study. *Fertil Steril* 2011; 95(3): 915–921.