Review Article

Access this article online
Quick Response Code:
Website:
www.jehp.net
DOI:
10.4103/jehp.jehp_1798_23

Machine learning in predicting infertility treatment success: A systematic literature review of techniques

Shirin Dehghan, Hamid Moghaddasi, Reza Rabiei, Hamid Choobineh[1], Keivan Maghooli[2], Mojtaba Vahidi-Asl[3]

Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [1]Department of Laboratory Sciences, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran, [2]Department of Biomedical Engineering Science and Research Branch Islamic Azad University, Tehran, Iran, [3]Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran

Address for correspondence:
Prof. Reza Rabiei, Department of Health Information Technology and Management, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Darband St., Qods Square (Tajrish), Tehran, Iran, Post Code: 1971653313.
Phone: 00989122195276
E-mail: R.Rabiei@sbmu.ac.ir

Received: 04-11-2023
Accepted: 22-01-2024
Published: 28-03-2025

Abstract:

Assisted reproductive technology (ART) is one of the major developments that has had a significant impact on infertility treatment. A predictive model of ART success based on machine learning (ML) techniques can provide a robust basis for estimating treatment success. This study aimed to identify predictive models of ART success and their determinants. A systematic search was conducted in PubMed, Web of Science, Scopus, and Embase. Data extraction involved collecting data in studies on dataset characteristics, ML techniques, and predictive model performance indicators. The search resulted in 3655 records, of which 27 papers were selected for analysis. ML publications in ART prediction have been in the past 5 years. In general, 107 various features were reported in all reviewed studies. Female age was the most common feature used in all identified studies. Most studies (96.3%) applied a supervised approach to develop predictive models. Among all, support vector machine (SVM) was the most frequently applied technique (44.44%). Nineteen different indicators have been used in studies to evaluate the model performance. 74.07% of the reviewed papers reported area under the receiver operating characteristic (ROC) curve (AUC) as their performance indicator. Accuracy (55.55%), sensitivity (40.74%), and specificity (25.92%) were also commonly reported. ML has the potential to bring hope to infertile couples and to facilitate making challenging decisions. Considering relevant contributing factors and ML techniques is critical for reliable predictive modeling.

Keywords:

Fertilization invitro, Fertility/therapy, Sperm injection, intracytoplasmic, Reproductive techniques, invitro, Machine learning

Introduction

Infertility, an emotionally stressful disability with a considerable social burden, is a growing public health challenge. Infertility is defined as the failure of a couple to conceive after 1 year of regular, unprotected intercourse.[1,2] Globally, it is estimated to affect 8–12% of couples of reproductive age, which constitutes 186 million people.[3]

Nowadays, more couples seek infertility treatment due to increased public awareness, convenient availability, and improvements in infertility treatments. Assisted reproductive technology (ART) is one of the major developments that has had a great impact on infertility treatment. It encompasses all techniques involving direct manipulation of oocytes outside of the body. The most common forms of ART are *in vitro* fertilization (IVF) and intracytoplasmic sperm injection (ICSI).[4] IVF comprises a sequence of steps beginning with ovarian stimulation and oocyte retrieval, fertilization, and embryo culture in the laboratory and embryo transfer to

How to cite this article: Dehghan S, Moghaddasi H, Rabiei R, Choobineh H, Maghooli K, Vahidi-Asl M. Machine learning in predicting infertility treatment success: A systematic literature review of techniques. J Edu Health Promot 2025;14:103.

the uterus for implantation. ICSI is very similar to IVF, with the difference that in this method during artificial insemination, sperm is injected into the cytoplasm of the oocyte.[2] So far, more than 8 million IVF children have been delivered alive worldwide.[5]

Although ART turns out to be a breakthrough in infertility treatment, it does not guarantee success. Compared to the pregnancy rate of 20% in healthy couples, the overall success rate has just been slightly higher and reaches 30–40%.[2,6] Also, there are complications associated with this treatment that could threaten both mother and child's health and life.[6-9] Besides the low rate of success and risk of adverse effects, ART is an expensive treatment procedure. Infertile couples undergo financial burdens because of the high costs of it.[10] This, in addition to the complexity of decision-making in medicine,[11,12] makes it crucial to uptake ART as a treatment for infertile couples. Infertility clinicians use patients' specific characteristics, together with their knowledge and clinical experience, to recommend this treatment when it is likely to succeed. Yet, due to the number of variables and complex relations among them, the prediction of infertility treatment success turns out to be a challenging task. The complexity could bias the clinician's judgment and affect clinical outcomes.[13,14]

Artificial intelligence has increasingly progressed over the past decade, and its applications have extended to various fields, including medicine. Machine learning (ML), a subset of artificial intelligence, enables systems to detect complex patterns in biomedical data and use them in line with clinical activities.[15,16] Systems based on ML techniques could train a classifier model with a large amount of medical data, which make it suitable for medical prediction.[16] In reproductive medicine, the ML techniques could help to improve the prediction of infertility treatment success resulting in making informed decisions for infertile couples. A predictive model of ART success based on ML techniques could provide a robust basis for the estimation of pregnancy chance by identifying interlinks among variables, making a realistic expectation for both clinicians and couples.

Predictive models can vary depending on the number and combination of features, size, and variety of the dataset and type of ML techniques. Accordingly, they could have different performances in terms of accuracy, area under receiver operating characteristic (AUC), sensitivity, specificity, etc., This variation highlights the importance of examining these elements in the context of infertility treatment. By thoroughly examining the previous studies, this review aims to contribute to a better understanding of the predictive modeling landscape in infertility treatment. Such insights increase our understanding of the field and guide future researchers and practitioners in making informed decisions about choosing algorithms, feature selection, and dataset characteristics to develop more robust and clinically relevant prediction models.

To the best of our knowledge, no systematic review has yet addressed this topic. This study aimed to identify predictive models of ART success and their determinants.

Materials and Methods

Study design

This study was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to report on evidence from the studies that were included in this systematic review.[17]

The questions examined in this review were as follows:
1. What features were used in predictive modeling for infertility success?
2. What ML techniques were applied in predictive modeling?
3. What performance indicators were reported in reviewed studies?

A systematic search was conducted in PubMed, Web of Science, Scopus, and Embase, in the time frame of 2000 to 2022, based on terms in the title, abstract, MeSH terms, and keywords. The final literature search was performed on November 18, 2022. The detailed search strategy is outlined in Figure 1. Different terms were applied for the two main concepts of "ART" and "ML." Box "A" contains all the terms on the "ART" concept, and box "B" contains all terms related to the concept of "ML." MeSH terms and Emtree terms are seen as highlighted ones. The search was performed by combining these two groups using the Boolean operator "AND." Truncation symbols, phrase searching, and other techniques were used to do a comprehensive and specific search.

Eligibility criteria

This review focused only on the full text of original papers published in English that addressed predictive models of ART (IVF/ICSI) outcomes based on an ML technique. Studies on the selection of embryos based on analyzing data of embryo images and predictive models based on any sort of deep learning approach were excluded.

Data extraction and synthesis

In the screening step, three authors (ShD, RR, and HM) examined the papers by their titles and abstracts and irrelevant studies were removed. In the eligibility step, the full text of the selected papers was independently

ART	ML
1. "Fertilization in vitro" 2. "Reproductive techniques, assisted" 3. "Sperm injections, intracytoplasmic" 4. "Embryo transfer" 5. "In vitro fertilization" 6. IVF 7. "In vitro fertility" 8. "Assisted reproduction" 9. "Assisted reproductive" 10. ICSI 11. "Intracytoplasmic sperm injections" 12. "Assisted conception" 13. "Huma-assisted reproduction" 14. reproduction medicine" 15. "Reproductive medicine" 16. "Reproductive technology" 17. "Reproduction technology" 18. "Reproductive techniques" 19. "Reproduction techniques" 20. "Human reproduction" 21. "Infertility" 22. "Test tube fertilization" 23. "Test tube baby" 24. "Assisted reproductive technology" 25. "Assisted reproductive techniques" 26. "Blastocyst transfer" 27. "Tubal embryo transfer" 28. "Tubal embryo stage transfer" 29. "Infertility therapy" 30. "Reproductive procedure" a	1. "Data mining" 2. "Artificial intelligence" 3. "Learning algorithm" 4. "Machine intelligence" 5. "Computational intelligence" 6. "Computer reasoning" 7. "Computer vision systems" 8. "Knowledge acquisition" 9. "Knowledge representation" 10. "Knowledge discovery" 11. "Learning machine" 12. "Learning system" 13. "Machine learning" b

Figure 1: Search strategy (Keywords related to (a) ART & (b) ML)

evaluated by two authors (ShD and RR) according to the inclusion or exclusion criteria. The third author (HM) was available to solve disagreements by brainstorming and consensus.

A qualitative and quantitative analysis of the studies was conducted. Data extraction involved collecting data in studies on dataset characteristics (number of records, number and types of features involved, and feature selection method), ML techniques (number and types of methods), and predictive model performance indicators. The quantitative analysis was conducted in Statistical Package for the Social Sciences (SPSS) version 26. The first author's name, publication year, and the study's location were also extracted. Disagreements were again solved by consensus involving the third author. The result of this analysis is adjusted in the form of a structured table.

Quality assessment

Two authors independently appraised the quality of the papers using the PROBAST checklist. PROBAST is a tool developed to evaluate the bias risk and applicability of prediction model studies.[18] It was created to give an organized and transparent technique for assessing the methodological quality of prediction model investigations. It has 20 signaling questions divided into four domains (participants, predictors, outcome, and analysis).

Ethical consideration

This study was approved by the Ethics Committee of Shahid Beheshti University of Medical Sciences (IR.SBMU. RETECH.REC.1400.695). The authors declare that no human subjects were involved in this research.

Result

As illustrated in Figure 2, searching databases resulted in 3655 records in total (PubMed: 802, Embase: 928, Scopus: 1281, and Web of Science: 644). A total of 1585 duplicates were removed using EndNote. The title and abstract of the remaining 2070 articles were screened, of which 106 eligible articles were identified for full-text assessment. In the next step, the authors reviewed the full text of the articles and excluded 79 articles, according to the inclusion or exclusion criteria. Studies were excluded if they did not apply a predictive model for IVF/ICSI outcome (27 records), did not use ML techniques in the predictive modeling (9 records), or did not include important study details (5 records). In addition, articles were excluded if there were conference papers, review articles, book chapters (26 records), and non-English papers (1 record) and if the full text was not available (11 records). The final set included 27 articles published from 2009 to 2023 [Table 1].

Dataset characteristics and ART features

A large amount of high-quality training data is required to obtain an accurate prediction model. The first step is to collect and store data appropriately. In predicting IVF/ICSI success, retrospective data are usually extracted from the medical records of infertile couples who visit an infertility treatment center. These data are reported as participants or treatment cycles in studies. Table 2 presents the dataset characteristics and features. Two studies were excluded as they reported the number of transferred embryos instead of treatment cycles or the number of participants. Among all the studies, only one study merged the data of multiple IVF centers across the UK.[26] The analysis applied to the data showed that the median number of participants in each study was 1048 and the mean number was 12723 (standard deviation (SD) =33542). The minimum number of participants was 62,[29] and the maximum was 141160.[26]

In less than half of the studies, feature selection methods were used to identify features with greater predictive power. The goal of feature selection methods is to help create more robust models to increase predictive performance. The feature selection methods used in the studies can be divided into three main categories: expert opinions, statistical methods, and ranking and ML techniques. In the first group, features are selected mainly based on subject knowledge, expert opinion, and clinical practice guidelines (CPGs). Only four studies used this method for feature selection.[19,24,25,45] In the second group, statistical methods are used to select features. Two studies used t-tests, correlation

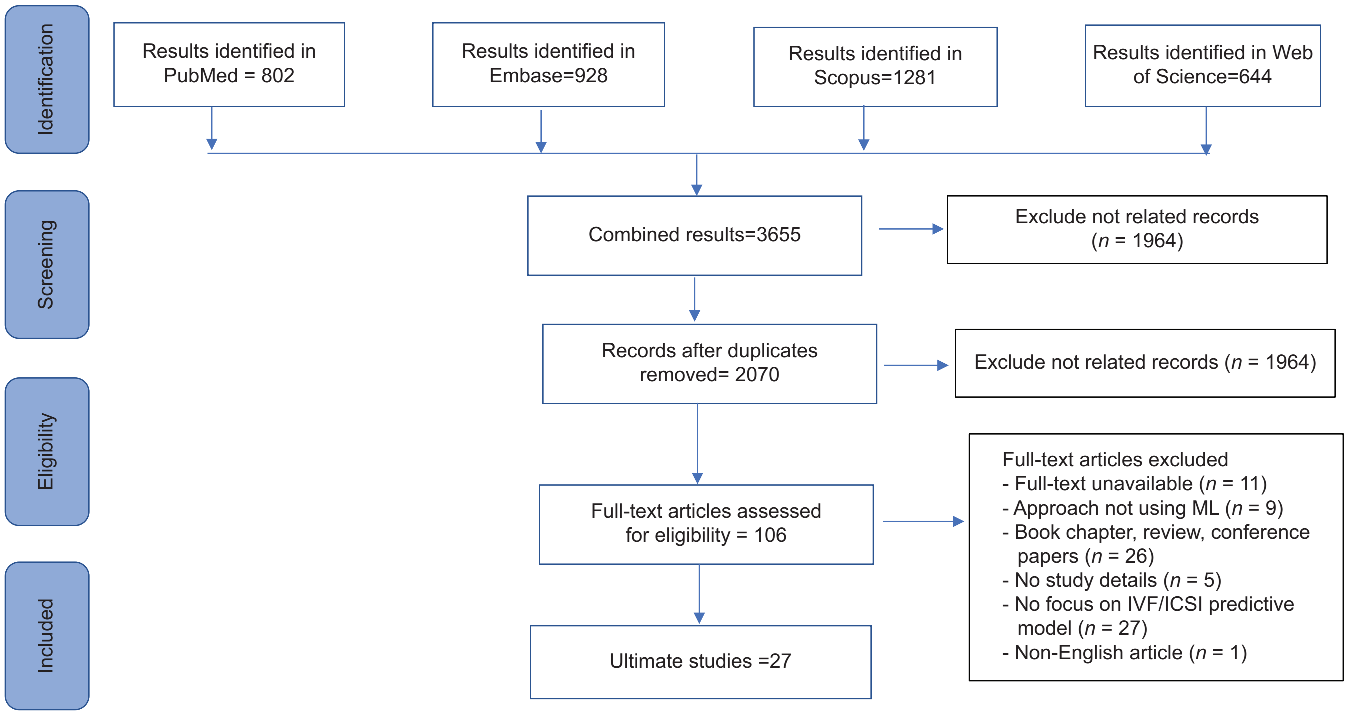

Figure 2: Study selection and data extraction

analysis, and multivariable logistic regression.[27,30] The third group used ranking algorithms (hill-climbing algorithm, principal component analysis method, information gain, and Johnson algorithm) and ML techniques (decision tree (DT), genetic algorithm, and random forest (RF)). Seven studies used these algorithms to define the features.[22,29,32,38,39,40,42] In eleven studies, no feature selection method was employed. Furthermore, three studies did not report any strategy for feature selection.

All studies except one[31] reported the features elaborately in a constructed table. The mean number of features, used in each study, was 21 (SD = 16). The minimum number of features was 6,[35] and the maximum was 78.[25] In general, 107 various features were reported in all reviewed studies [Table 3]. The features used in the studies could be classified into eight main categories: demographics (3 features), medical history (19 features), laboratory findings (24 features), ovary stimulation (10 features), oocyte (15 features), sperm (13 features), and embryo (23 features). Female age was the most common feature used in all of the identified studies. Follicle-stimulating hormone (FSH) was the second most frequently used feature in 19 studies (73.07%). Other than women's age and FSH, the cause of infertility and duration of infertility were repeated in 13 (50%) and 12 (46.15%) studies, respectively. Estradiol, endometrial thickness, and body mass index (BMI) as three common features were used in 11 studies (42.30%). Fifty-five (51.4%) features were used only once in studies.

ML techniques in predicting ART success

The mean number of ML techniques employed in each study was 3 (SD = 2). Most studies (96.3%) applied a supervised approach to develop predictive models, and only one study applied an unsupervised approach to model the data.

A wide range of different types and combinations of ML techniques have been used in the studies. Among all, support vector machine (SVM) was the most frequently applied technique (44.44%). Both RF and artificial neural network (ANN) were applied in 40.74% of studies. DT and naïve Bayes (NB) were applied in 37.3% and 29.62% of studies, respectively. The number of studies applying ML techniques to IVF or ICSI prediction has grown during the past decade. Figure 3 displays the trend of applied techniques in the studies. The top most-frequent ML techniques were compared together.

Model performance for ART outcome prediction

There was considerable variety in the way authors chose to evaluate the predictive model performance. Nineteen different indicators have been used in studies to evaluate the model performance. Of the 24 studies included in the review, 20 studies (74.07%) reported AUC as their performance indicator. Accuracy (55.55%), sensitivity (40.74%), and specificity (25.92%) were also commonly reported. Seventeen percent of studies reported positive predictive value (PPV), while negative predictive value (NPV), F-measure, recall, and false

Table 1: Details of studies identified

Reference	ML techniques	Dataset	Performance indicators						
			Accuracy	Sensitivity	specificity	AUC	Precision	PPV	NPV
Tian, *et al.* 2023[19]	Bayesian network model	106,640 cycles, 24 features	91.3	-	-	0.997	-	-	-
Wang, *et al.* 2022[20]	Comparing RF/LR	24,730 records, 9 features	64.78	66.58	64.16	0.7208	-	-	-
Mehrjerd, *et al.* 2022[21]	Comparing LR, RF, KNN, ANN, SVM, GNB	1931 records, 14 features	-	0.76	-	-	-	0.80	-
Liang, *et al.* 2021[22]	RF	2189 cycles, 28 features	76.9	0.60%	0.91%	0.83	-	-	-
Kozar, *et al.* 2021[23]	Comparing SVM/NB/RF/MLP	1029 cycles, 24 predictors	-	0.432	0.756	0.66	-	-	-
Amini, *et al.* 2021[24]	Comparing LR/SVM/XGBoost/RF/NB/LDA	6071 cycles, 15 features	-	-	-	0.60	-	-	-
Raef, *et al.* 2020[25]	Comparing NB/KNN/ANN/RF/DT/SVM	1360 transferred embryos, 78 features	90.4	90.36	90.44	93.74	-	90.36	-
Goyal, *et al.* 2020[26]	Comparing LR/KNN/MLP/DT/1-DNN/RF/AdaBoost	141,160 records, 25 features	-	-	-	84.60%	77%	-	-
Vogiatzi, *et al.* 2019[27]	ANN	426 cycles, 12 features	69.19	69.2	69.19	-	-	36.96	89.61
Qiu, *et al.* 2019[28]	RF/XGBoost/SVM/LR	7188records, 8 features	0.70	-	-	0.73	-	-	-
Mostaar, *et al.* 2019[29]	MLP	62 records, 18 features	-	-	-	0.9394-0.9990	-	-	-
Zhang, *et al.* 2019[30]	Comparing clustering/SVM/C-SVM	11,190 records, 11 features	-	-	-	0.70	-	-	-
Source	**ML techniques**	**Dataset**	**Performance indicators**						
			Accuracy	**Sensitivity**	**specificity**	**AUC**	**Precision**	**PPV**	**NPV**
Blank, *et al.* 2019[31]	Comparing RF/MLR	1052 records, 32 features	-	0.84	-	0.74	-	-	-
Hassan, *et al.* 2018[32]	MLP/SVM/CART/RF/C4.5	1048 records, 25 features	97.42	-	-	0.973	-	-	-
Hafiz, *et al.* 2017[33]	SVM/RPART/RF/AdaBoost/1NN	486 records, 29 features	83.96	-	98.03	84.23	-	90.14	-
Siristatidis, *et al.* 2016[34]	ANN	300 records, 12 features	-	-	-	-	-	> 75%	> 75%
Milewska, *et al.* 2016[35]	DT	610 cycles, 6 features	-	-	-	66-68	-	-	-
Milewska, *et al.* 2015[36]	Discriminant analysis	610 cycles, 18 features	-	0.512	-	0.73	-	-	-
Güvenir, *et al.* 2015[37]	Ranking algorithm SERA	1456 records, 64 features	0.844	-	-	0.833	-	-	-
Uyar, *et al.* 2015[38]	Comparing NB/SVM/DT/MLP/RBF network/KNN	2453 records, 18 features	80.4%	63.7%	-	0.754	-	-	-
Durairaj, *et al.* 2014[39]	ANN	250 records, 9 features	90%	-	-	-	-	-	-
Kakhki, *et al.* 2013[40]	Comparing NB, Bayes-N, MP-Bayes, Greedy	942 records, 13 features	92.4%	82%	97%	-	-	-	-
Gianaroli, *et al.* 2013[41]	Comparing DT/BN	388 cycles, 8 features	81.5%	-	-	0.72	-	-	-
Guh, *et al.* 2011[42]	DT	5275 patients, 38 features	73.2%	-	-	-	-	-	-

Contd...

Table 1: Contd...

Reference	ML techniques	Dataset	Performance indicators						
			Accuracy	Sensitivity	specificity	AUC	Precision	PPV	NPV
Nanni, *et al.* 2011[43]	SVM	98 cycles, 25 features	-	-	-	0.79	-	-	-
Uyar, *et al.* 2010[44]	Comparing NB/KNN/DT/SVM/MLP/RBF	2453 patients, 18 features	-	67%	-	0.739	-	-	-
Uyar, *et al.* 2009[45]	SVM	546 records, 17 features	82.7%	-	-	-	-	-	-

RF: random forest, SVM: support vector machine, NB: naïve Bayes, MLP: multi-layer perceptron, LR: logistic regression, LDA: liner discriminant analysis, KNN: k-nearest neighbor, ANN: artificial neural network, DT: decision tree, MLR: multivariate logistic regression, RPART: recursive partitioning, RBF: radial basis network, BN: Bayesian network, LGBM: light gradient boosting machine, GNB: gradient naïve Bayes

Table 2: Details of datasets in reviewed studies

Source	Number of participants	Location	Feature selection method	No. of features
Tian, *et al.* 2023[19]	106640 cycles	Peking, China	Previous studies, expert opinion	17
Wang, *et al.* 2022[20]	24730 cycles	Taipei Medical University Hospital, Taiwan	-	9
Mehrjerd, *et al.* 2022[21]	1931 cycles	Mashhad, Iran	-	13
Liang, *et al.* 2021[22]	2189 cycles	Peking University, China	Genetic algorithm	9
Kozar, *et al.* 2021[23]	1029 cycles	University of Medical Centre Maribor, Slovenia	-	24
Amini, *et al.* 2021[24]	6071 cycles	Royan Institute, Tehran	-	15
Raef, *et al.* 2020[25]	500 patients, 1360 transferred embryos	Tabriz, Iran	Expert opinion, guidelines, SVM-FS technique	78
Goyal, *et al.* 2020[26]	141,160 records	IVF centers across UK	Linear support vector classifier (linear SVC), tree-based feature selection	25
Vogiatzi, *et al.* 2019[27]	257 patients, 426 cycles	Attikon University, Greece	*t*-test, correlation	12
Qui, *et al.* 2019[28]	7188 patients	Shengjing Hospital, China	Subject knowledge, previous studies, and guidelines	8
Mostaar, *et al.* 2019[29]	62 patients	Ayatollah Taleghani Hospital, Tehran	PCA	18
Zhang, *et al.* 2019[30]	11,190 patients	Huazhong University of Science and Technology, Wuhan, China	MLR	11
Blank, *et al.* 2019[31]	1052 records,	Ghent University Hospital, Belgium	-	32
Hassan, *et al.* 2018[32]	1048 patients	Istanbul, Turkey	Hill-climbing algorithm	25
Hafiz, *et al.* 2017[33]	486 patients	Shiraz, Iran	-	29
Siristatidis, *et al.* 2016[34]	>300 patients	University of Athens, Greece	-	12
Milewska, *et al.* 2016[35]	610 cycles	Acacio Fertility Center, USA	-	6
Source	**Dataset records**	**Location**	**Feature selection method**	**No. of features**
Milewska, *et al.* 2015[36]	610 IVF cycles	USA	-	18
Güvenir, *et al.* 2015[37]	1456 patients	Ankara, Turkey	-	64
Uyar, *et al.* 2015[38]	2453 embryo	Turkey, Istanbul	Information gain	18
Durairaj, *et al.* 2014[39]	250 patients	Tamil Nadu, India	Johnson algorithm	9
Kakhki, *et al.* 2013[40]	942 records	Mashhad, Iran	-	13
Gianaroli, *et al.* 2013[41]	388 cycles	Lugano, Switzerland	-	8
Guh, *et al.* 2011[42]	5275 patients	Los Angeles, USA	Genetic algorithm	38
Nanni, *et al.* 2011[43]	98 cycles	Università di Bologna, Italy	-	21
Uyar, *et al.* 2010[44]	2453 patient	German Hospital, Istanbul, Turkey	-	18
Uyar, *et al.* 2009[45]	546 embryo records	Turkey, Istanbul	Embryologist's opinion	17

alarm ratio were all reported by 8%. Other indicators were used only in one study.

Fourteen studies have compared the performance of multiple ML techniques to choose the best one, based on performance indicators. In eight of these studies, RF performed the best. Table 4 presents the comparison among the six most frequent techniques in pairs. The most common comparison was between RF and SVM. RF outperformed in 83% of the cases (five of six studies). Five studies compared RF and ANN, in which RF surpassed all of them. Also, RF compared with NB in four studies and performed better in all of them. Two studies compared RF and DT, which RF outperformed.

Table 3: Features in reviewed studies

Category	Name	Count	Percent
Demographic	Female age	26	100%
	Male age	6	23.07%
	Age group	1	4.3%
Clinical history	Duration of infertility	12	46.15%
	Cause of infertility	13	50%
	BMI	11	42.30%
	Type of infertility	10	38.46%
	No. of previous IUI/IVF/ICSI	6	26.1%
	Previous live birth	4	15.38%
	Surgical history	3	11.54%
	Smoking	3	11.54%
	Previous abortion	3	11.54%
	History of abortion	2	7.70%
	Total number of IVF pregnancies	2	7.70%
	Medical disease history	2	7.70%
	Gravidity	2	7.70%
	Previous miscarriage	2	7.70%
	Age at menarche	1	3.85%
	Menarche >12 years	1	3.85%
	Total no. of IVF live birth	1	3.85%
	Contraception duration	1	3.85%
	Female blood type	1	3.85%
	Follicle-stimulating hormone (FSH)	19	78.3%
Laboratory findings	Estradiol	11	42.30%
	Endometrial thickness	11	42.30%
	Antral follicle count (AFC)	10	38.46%
	Anti-Mullerian hormone (AMH)	6	26.1%
	No. of follicles >17 mm	6	26.1%
	luteinizing hormone	5	19.23%
	Polycystic ovarian syndrome (PCOs)	5	19.23%
	Endometriosis	4	15.38%
	HOMA-IR	1	3.85%
	Perimenopause	1	3.85%
	Dyspareunia	1	3.85%
	TSH	1	3.85%
	Diminished ovarian reserve	1	3.85%
	Intrauterine adhesion	1	3.85%
	Hysteroscopic parameter findings	1	3.85%
	Vitamin 3 deficiency	1	3.85%
	PGD test	1	3.85%
	Endometrial vascularization index	1	3.85%
	Endometrial flow index	1	3.85%
	Endometrial VFI	1	3.85%
	Subendometrial VI	1	3.85%
	Subendometrial VFI	1	3.85%
	Volume of endometrial	1	3.85%
Ovary stimulation	No. of retrieved oocytes	10	38.46%
	Total gonadotropin	4	15.38%
	Duration of stimulation	4	15.38%
	Day of trigger	1	3.85%
	Stimulation type	1	3.85%
	Antagonist	1	3.85%
	Ovulatory stimulants	1	3.85%
	Surgical physician	1	3.85%
	Catheter code	1	3.85%

Contd...

Table 3: Contd...

Category	Name	Count	Percent
	Oocyte retrieving technique	1	3.85%
Oocyte	No. of MII quality oocytes	4	15.38%
	No. of follicles >14 mm	3	13%
	No. of GV quality oocytes	3	13%
	No. of fertilized oocytes	2	7.70%
	Frozen or fresh oocyte	2	7.70%
	Fertilization rate	2	7.70%
	Max follicle size	1	3.85%
	Avg. follicle size	1	3.85%
	No. of follicles 14<>17 mm	1	3.85%
	Eggs thawed	1	3.85%
	Fresh eggs collected	1	3.85%
	Maturity of retrieved oocytes	1	3.85%
	Uterus depth	1	3.85%
	No. of MI quality oocytes	1	3.85%
	No. of nonviable oocytes	1	3.85%
	Total sperm count	6	23.07%
Sperm	Sperm quality grade	5	21.8%
	Sperm motility	5	19.23%
	Motile sperm concentration	4	15.38%
	Sperm morphology	4	15.38%
	Sperm concentration	2	7.70%
	Frozen or fresh sperm	2	7.70%
	Method of sperm collection	2	7.70%
	Sperm injection volume	2	7.70%
	Ejaculate volume	1	3.85%
	Spermogram	1	3.85%
	Insemination technique	1	3.85%
	Sperm progress	1	3.85%
Embryo	The number of embryo transfers	11	42.30%
	Type of cycle (IVF/ICSI/IUI)	9	30.4%
	Total number of embryos	8	34.8%
	Embryo transfer day	7	30.4%
	Fresh or frozen cycle	6	23.07%
	Quality of embryo	4	15.38%
	Early cleavage morphology	3	11.54%
	Nucleus characteristics	3	11.54%
	Fragmentation rate	3	11.54%
	Appearance of cytoplasm	3	11.54%
	Thickness of zona pellucida	3	11.54%
	Early cleavage inspection time	2	7.70%
	Difficulty during ET	1	3.85%
	ET/2PN	1	3.85%
	Top quality embryo	1	3.85%
	TQE D3/2PN	1	3.85%
	Embryo grade	1	3.85%
	No. blastomere day 5	1	3.85%
	No. blastomere day 6	1	3.85%
	Transferring physicians	1	3.85%
	Chamotte	1	3.85%
	No. of cleavage	1	3.85%
	No. of injected oocytes	1	3.85%

NB and SVM were compared in two studies, while NB demonstrated superior performance in both of them. NB also outperformed ANN in two studies. One study compared SVM and ANN, while SVM showed superior performance. SVM performed better in comparison with DT in another one. Only one study compared XGBoost

Table 4: Comparing technique performance: Counts of studies where technique outperformed technique B

Inferior performance	Superior performance					Decision tree	Total
	Random forest	Naïve base	Support vector machine	XGBoost	Neural network		
Random forest	-	0	1	1	0	0	2
Naïve base	4	-	0	0	0	0	3
Support vector machine	5	2	-	1	0	0	7
XGBoost	0	0	0	-	0	0	0
Neural network	5	2	1	0	-	0	7
Decision tree	2	3	1	0	0	-	6
Total	16	7	3	2	0	0	

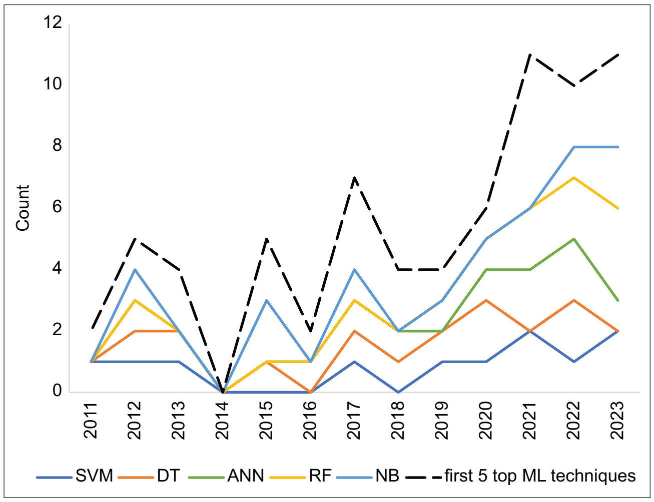

Figure 3: ML publication and applied techniques by year

and RF, in which XGBoost performed better. Also, XGBoost presented superior performance rather than SVM in another study.

Discussion

This review was performed to review the application of ML techniques to predict infertility treatment success. Of the 3655 records retrieved through searching four literature databases, 27 eligible studies matched the inclusion criteria. These studies mainly focused on developing ML predictive models for the ART outcome.

Results showed that the number of ML publications in ART prediction has increased over the past 5 years. Since 2017, the number of publications predicting the outcome of infertility treatment has grown substantially. The interest in using ML to predict ART outcomes has increased in recent years, but still there is substantial scope for further applications. These findings are consistent with those of Dhombres *et al*.,[46] who observed a significant rising trend in AI contribution in obstetrics and gynecology, a wide range of AI approaches applied across all the subdomains.

In this review, female age was identified as the most significant feature, as it was reported in all reviewed studies as well. This was consistent with Meena and Vijayalakshmi's research findings, as they found female age to be the most significant factor influencing female fertility through statistical analysis and data mining techniques. However, Meena and Vijayalakshmi[47] found BMI and thyroid-stimulating hormone (TSH) levels as other influencing factors, which contradicts the findings of this review. Curchoe and Bormann[48] pointed out that not a specific set of features has been determined as the most contributing factors in successful infertility treatment till now. This could be known and elucidated through the use of artificial intelligence. The findings showed that the studies used 107 various features in predictive models. The plentifulness of factors affecting the outcome of ART indicates the usefulness of ML methods for prediction.[49] This is in line with Vaughan *et al*.'s[13] research findings as well, who concluded that the abundance of features related to infertility treatment is well suited for the application of ML.

The results showed that eleven studies did not use the feature selection method and three others did not report whether they used any feature selection or not. The robustness of an ML model depends on the contribution of predicting features in the model. The accuracy of a predictive model mainly depends on how influential variables are incorporated. "Feature selection method" as a crucial step in data preprocessing can identify the most significant features needed to improve the performance.[50,51] Besides, overfitting is an event that leads to poor model performance and is associated with applying too many irrelevant features in the model. Overfitting occurs when a predictive model adapts to the details of the training dataset and affects the model performance on new data.[52] Using the feature selection method to identify the most significant variables could overcome overfitting.

The vast majority of the studies have applied supervised ML approaches to predict success. In the supervised learning approach, an ML technique uses a labeled dataset to develop a predictive model. Then, the predictive model maps the input data to predefined groups or classes. As classes were defined before examining the data, it is referred to as supervised learning.[14] Results revealed that supervised ML approaches could be suitable for the prediction of infertility treatment success. This is in agreement with the findings of Wang *et al*.[16] who reported the successful application of supervised learning methods to the prediction of ART success.

Although nearly all studies used supervised ML techniques, there was significant heterogeneity in the applied modeling approaches. Studies differed concerning the ML techniques and features used for predictive modeling for infertility. The most common technique reported in studies was SVM, followed by RF, ANN, DT, and NB. In this review, ML techniques were compared to performance indicators in 14 studies. According to findings, in more than half of these studies, RF outperforms other ML techniques in supervised learning tasks, while the ensemble techniques are less vulnerable to overfitting[53] and are recognized as superior to individual classifiers in prediction performance.[54] Most studies reported AUC, accuracy, sensitivity, and specificity as performance indicators. Although further metrics are required for better predictive performance, only a limited number of studies have considered PPV and NPV, despite their critical role in the performance prediction assessment.[55]

Limitation and recommendation

There are limitations to this review, which can be considered in future studies. First, despite our efforts to perform a comprehensive search across multiple online databases, we may have missed non-English studies reporting ML predictive models.

This review may likely be influenced by publication bias. Positive results are typically more likely to be published.[56] Selective outcome reporting may have affected our findings since we found no studies reporting failed predictive models. Nevertheless, in this review, an attempt was made to report and compare the performance of all the different predictive models in the studies.

This review highlights the need for continued research and innovation in the application of artificial intelligence in reproductive medicine. Also, it provides a snapshot of the current landscape in this field. The combination of features, along with a thorough understanding of ML techniques, is identified as two critical elements for the development of predictive ART models. Considering this will have a major impact on this realm, as technology and data availability continue to advance.

Conclusion

In this study, we aimed to gain insight into the application of ML techniques to predict infertility treatment success. To the best of our knowledge, this is the first review that systematically identifies predictive models of ART outcome success and their determinants. In this review, female age was identified as the most significant feature in studies. In the majority of reviewed studies, supervised ML approaches have been used to predict success. The most common technique used in the studies was SVM, followed by RF, ANN, DT, and simple Bayes. This systematic review provides insight into the evolving landscape of ML applications in predicting ART success and guiding future researchers and practitioners in developing more robust and clinically relevant prediction models. The synthesis of current knowledge underscores the importance of standardized methodologies and understanding the factors influencing ART outcomes for informed decision-making in infertility treatment.

Acknowledgement

The authors declare that there are no conflicts of interest to influence the work reported in this paper. No funding sources were used in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Seyedi SS, Hosseini A, Rabiei R, Asadi F, Moghaddasi H. Infertility information system with an approach to data architecture: A systematic review. Am J Biomed Sci Res 2019;5:254-61.
2. Taylor HS, Pal L, Sell E. Speroff's Clinical Gynecologic Endocrinology and Infertility. Philadelphia, USA, Lippincott Williams and Wilkins; 2019.
3. Vander Borght M, Wyns C. Fertility and infertility: Definition and epidemiology. Clin Biochem 2018;62:2-10.
4. Casanova R, Chuang A, Goepfert AR, Hueppchen NA, Weiss PM. Beckmann and Ling's Obstetrics and Gynecology. Netherlands, Wolters Kluwer; 2019.
5. Fauser BCJM. Towards the global coverage of a unified registry of IVF outcomes. Reprod BioMed Online 2019;38:133-7.
6. Berek JS. Berek and Novak's Gynecology. Netherlands, Wolters Kluwer; 2020.
7. Kuivasaari-Pirinen P, Raatikainen K, Hippeläinen M, Heinonen S. Adverse outcomes of IVF/ICSI pregnancies vary depending on aetiology of infertility. ISRN Obstet Gynecol 2012;2012:451915. doi: 10.5402/2012/451915.
8. Nagata C, Yang L, Yamamoto-Hanada K, Mezawa H, Ayabe T, Ishizuka K, *et al*. Complications and adverse outcomes in pregnancy and childbirth among women who conceived by assisted reproductive technologies: A nationwide birth cohort study of Japan environment and children's study. BMC Pregnancy Childbirth 2019;19:77.
9. Hilbert SM, Gunderson S. Complications of assisted reproductive technology. Emerg Med Clin North Am 2019;37:239-49.
10. Inhorn MC, Patrizio P. Infertility around the globe: New thinking on gender, reproductive technologies and global movements in the 21st century. Hum Reprod Update 2015;21:411-26.
11. Afrash MR, Rabiei R, Hosseini A, Salari S, Sepehri M. Impact of computerized provider order entry on chemotherapy medication

errors: A systematic review. 2022;15:e120300. doi: 10.5812/ijcm-120300.

12. Bigdeli S, Baradaran HR, Ghanavati S, Arabshahi SKS. A qualitative approach to identify clinical uncertainty in practicing physicians and clinical residents. J Educ Health Promot 2022;11:278.
13. Vaughan D, Pan W, Yacoby Y, Seidler EA, Leung AQ, Doshi-Velez F, *et al*. The application of machine learning methods to evaluate predictors of live birth in programmed thaw cycles. Fertil Steril 2019;112:e273.
14. Ranjini K, Suruliandi A, Raja SP. Machine learning techniques for assisted reproductive technology: A review. J Circuits Syst Comput 2020;29:2030010. doi: 10.1142/S021812662030010X.
15. Rabiei R, Ayyoubzadeh SM, Sohrabei S, Esmaeili M, Atashi A. Prediction of breast cancer using machine learning approaches. J Biomed Phys Eng 2022;12:297-308.
16. Wang RJ, Pan W, Jin L, Li YH, Geng YD, Gao C, *et al*. Artificial intelligence in reproductive medicine. Reproduction 2019;158:R139-54.
17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Syst Rev 2021;10:89.
18. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, *et al*. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Ann Intern Med 2019;170:W1-33.
19. Tian T, Kong F, Yang R, Long XY, Chen LX, Li M, *et al*. A Bayesian network model for prediction of low or failed fertilization in assisted reproductive technology based on a large clinical real-world data. Reprod Biol Endocrinol 2023;21:8. doi: 10.1186/s12958-023-01065-x.
20. Wang CW, Kuo CY, Chen CH, Hsieh YH, Su EC. Predicting clinical pregnancy using clinical features and machine learning algorithms in *in vitro* fertilization. PLoS One 2022;17:e0267554. doi: 10.1371/journal.pone. 0267554.
21. Mehrjerd A, Rezaei H, Eslami S, Ratna MB, Ghaebi NK. Internal validation and comparison of predictive models to determine success rate of infertility treatments: A retrospective study of 2485 cycles. Sci Rep 2022;12:7216. doi: 10.1038/s41598-022-10902-9.
22. Liang R, An J, Zheng Y, Li J, Wang Y, Jia Y, *et al*. Predicting and improving the probability of live birth for women undergoing frozen-thawed embryo transfer: A data-driven estimation and simulation model. Comput Methods Programs Biomed 2021;198:105780. doi: 10.1016/j.cmpb. 2020.105780.
23. Kozar N, Kovač V, Reljič M. Can methods of artificial intelligence aid in optimizing patient selection in patients undergoing intrauterine inseminations? J Assist Reprod Genet 2021;38:1665-73.
24. Amini P, Ramezanali F, Parchehbaf-Kashani M, Maroufizadeh S, Omani-Samani R, Ghaheri A. Factors associated with *in vitro* fertilization live birth outcome: A comparison of different classification methods. Int J Fertil Steril 2021;15:128-34.
25. Raef B, Maleki M, Ferdousi R. Computational prediction of implantation outcome after embryo transfer. Health Inform J 2020;26:1810-26.
26. Goyal A, Kuchana M, Ayyagari KPR. Machine learning predicts live-birth occurrence before in-vitro fertilization treatment. Sci Rep 2020;10:20925. doi: 10.1038/s41598-020-76928-z.
27. Vogiatzi P, Pouliakis A, Siristatidis C. An artificial neural network for the prediction of assisted reproduction outcome. J Assist Reprod Genet 2019;36:1441-8.
28. Qiu J, Li P, Dong M, Xin X, Tan J. Personalized prediction of live birth prior to the first *in vitro* fertilization treatment: A machine learning method. J Transl Med 2019;17317.
29. Mostaar A, Sattari MR, Hosseini S, Deevband MR. Use of artificial neural networks and pca to predict results of infertility treatment in the icsi method. J Biomed Phys Eng 2019;9:679-86.
30. Zhang B, Cui Y, Wang M, Li J, Jin L, Wu D. *In vitro* fertilization (IVF) cumulative pregnancy rate prediction from basic patient characteristics. IEEE Access 2019;7:130460-7.
31. Blank C, Wildeboer RR, DeCroo I, Tilleman K, Weyers B, de Sutter P, *et al*. Prediction of implantation after blastocyst transfer in *in vitro* fertilization: A machine-learning perspective. Fertil Steril 2019;111:318-26.
32. Hassan MR, Al-Insaif S, Hossain MI, Kamruzzaman J. A machine learning approach for prediction of pregnancy outcome following IVF treatment. Neural Comput Appl 2018;32:2283-97.
33. Hafiz P, Nematollahi M, Boostani R, Jahromi BN. Predicting implantation outcome of *in vitro* fertilization and intracytoplasmic sperm injection using data mining techniques. Int J Fertil Steril 2017;11:184-90.
34. Siristatidis C, Vogiatzi P, Pouliakis A, Trivella M, Papantoniou N, Bettocchi S. Predicting IVF outcome: A proposed web-based system using artificial intelligence. *In vivo* (Athens, Greece) 2016;30:507-12.
35. Milewska AJ, Jankowska D, Cwalina U, Citko D, Więsak T, Acacio B, *et al*. Prediction of infertility treatment outcomes using classification trees. Stud. Log. Gramm. Rhetor. 2016;47:7-19.
36. Milewska AJ, Jankowska D, Cwalina U, Citko D, Więsak T, Acacio B, *et al*. Significance of discriminant analysis in prediction of pregnancy in IVF treatment. Stud Log Gramm Rhetor 2015;43:7-20.
37. Güvenir HA, Misirli G, Dilbaz S, Ozdegirmenci O, Demir B, Dilbaz B. Estimating the chance of success in IVF treatment using a ranking algorithm. Med Biol Eng Comput 2015;53:911-20.
38. Uyar A, Bener A, Ciray HN. Predictive modeling of implantation outcome in an *in vitro* fertilization setting: An application of machine learning methods. Med Decis Making 2015;35:714-25.
39. Durairaj M, Nandhakumar R, Ieee. An integrated methodology of artificial neural network and rough set theory for analyzing IVF data. 2014 International Conference on Intelligent Computing Applications, Coimbatore, 2014. p. 126-9.
40. Kakhki SA, Malekara B, Quchani SR, Khadem N. A model based on Bayesian network for prediction of IVF success rate. International Symposium on Advances in Science and Technology 7-8 March 2013. Bandar-Abbas, Iran: SASTech; 2013.
41. Gianaroli L, Magli MC, Gambardella L, Giusti A, Grugnetti C, Corani G. Objective way to support embryo transfer: A probabilistic decision. Hum Reprod 2013;28:1210-20.
42. Guh RS, Wu TCJ, Weng SP. Integrating genetic algorithm and decision tree learning for assistance in predicting *in vitro* fertilization outcomes. Expert Syst Appl 2011;38:4437-49.
43. Nanni L, Lumini A, Manna C. Artificial intelligence techniques for assisting the decision of making or postponing the embryo transfer. Perspectives on Pattern Recognition. Nova Science Publishers Inc, Hauppauge, NewYork 2011. p. 137-52.
44. Uyar A, Bener A, Ciray HN, Bahceci M. ROC based evaluation and comparison of classifiers for IVF implantation prediction. International Conference on Electronic Healthcare 2009 Sep 23 (pp. 108-111). Berlin, Heidelberg: Springer Berlin Heidelberg.
45. Uyar A, Ciray HN, Bener A, Bahceci M. 3P: Personalized pregnancy prediction in IVF treatment process. Electronic Healthcare: First International Conference, eHealth 2008, London, UK, September 8-9, 2008. Revised Selected Papers 1 2009 (pp. 58-65). Springer Berlin Heidelberg
46. Dhombres F, Bonnard J, Bailly K, Maurice P, Papageorghiou AT, Jouannic J-M. Contributions of artificial intelligence reported in obstetrics and gynecology journals: Systematic review. J Med Internet Res 2022;24:e35465. doi: 10.2196/35465.
47. Meena K, Vijayalakshmi N. Analysis of factors causing infertility in women using statistical analysis and association rule mining. Indian J Public Health Res Dev 2015;6:112-7.

48. Curchoe CL, Bormann CL. Artificial intelligence and machine learning for human reproduction and embryology presented at ASRM and ESHRE 2018. J Assist Reprod Genet 2019;36:591-600.
49. Narasimhan G, Victor A. Analysis of computational intelligence approaches for predicting disease severity in humans: Challenges and research guidelines. J Educ Health Promot 2023;12:334.
50. Swain J, VerMilyea MT, Meseguer M, Ezcurra D, Letterie G, Sánchez P, *et al*. AI in the treatment of fertility: Key considerations. J Assist Reprod Genet 2020;37:2817-24.
51. Fernandez EI, Ferreira AS, Cecílio MHM, Chéles DS, de Souza RCM, Nogueira MFG, *et al*. Artificial intelligence in the IVF laboratory: Overview through the application of different types of algorithms for the classification of reproductive data. J Assist Reprod Genet 2020;37:2359-76.
52. Jain V, Chatterjee JM. Machine Learning with Health Care Perspective: Machine Learning and Healthcare. Springer International Publishing; Germany, 2020.
53. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. Med 2021;2:642-65.
54. Pintelas P, Livieris IE. Special issue on ensemble learning and applications. Algorithms. 2020;13:140.
55. Trevethan R. Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017;5:307. doi: 10.3389/fpubh.2017.00307.
56. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias — An updated review. PLoS One 2013;8:e66844. doi: 10.1371/journal.pone. 0066844.