Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data

Eleonora Inácio Fernandez¹ · André Satoshi Ferreira¹ · Matheus Henrique Miquelão Cecílio¹ · Dóris Spinosa Chéles^{1,2} · Rebeca Colauto Milanezi de Souza¹ · Marcelo Fábio Gouveia Nogueira² · José Celso Rocha^{1,3}

Received: 8 March 2020 / Accepted: 3 July 2020 / Published online: 11 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Over the past years, the assisted reproductive technologies (ARTs) have been accompanied by constant innovations. For instance, intracytoplasmic sperm injection (ICSI), time-lapse monitoring of the embryonic morphokinetics, and PGS are innovative techniques that increased the success of the ART. In the same trend, the use of artificial intelligence (AI) techniques is being intensively researched whether in the embryo or spermatozoa selection. Despite several studies already published, the use of AI within assisted reproduction clinics is not yet a reality. This is largely due to the different AI techniques that are being proposed to be used in the daily routine of the clinics, which causes some uncertainty in their use. To shed light on this complex scenario, this review briefly describes some of the most frequently used AI algorithms, their functionalities, and their potential use. Several databases were analyzed in search of articles where applied artificial intelligence algorithms were used on reproductive data. Our focus was on the classification of embryonic cells and semen samples. Of a total of 124 articles analyzed, 32 were selected for this review. From the proposed algorithms, most have achieved a satisfactory precision, demonstrating the potential of a wide range of AI techniques. However, the evaluation of these studies suggests the need for more standardized research to validate the proposed models and their algorithms. Routine use of AI in assisted reproduction clinics is just a matter of time. However, the choice of AI technique to be used is supported by a better understanding of the principles subjacent to each technique, that is, its robustness, pros, and cons. We provide some current (although incipient) and potential uses of AI on the clinic routine, discussing how accurate and friendly it could be. Finally, we propose some standards for AI research on the selection of the embryo to be transferred and other future hints. For us, the imminence of its use is evident, providing a revolutionary milestone that will impact the ART.

Keywords Artificial intelligence · Assisted reproductive technologies · Prediction models · Embryo classification · Multilayer perceptron · Deep learning

✉ José Celso Rocha
jose.celso@unesp.br

Eleonora Inácio Fernandez
inacio.eleonora@gmail.com

André Satoshi Ferreira
andresatoshi.f@gmail.com

Matheus Henrique Miquelão Cecílio
mhmcilio@gmail.com

Dóris Spinosa Chéles
dorischeles2@gmail.com

Rebeca Colauto Milanezi de Souza
milanezi.souza@gmail.com

Marcelo Fábio Gouveia Nogueira
marcelo.fabio@unesp.br

¹ Laboratory of Applied Mathematics, Department of Biological Sciences, São Paulo State University (UNESP), Campus Assis, Av. Dom Antônio, São Paulo 2100, Brazil

² Laboratory of Embryonic Micromanipulation, Department of Biological Sciences, São Paulo State University (UNESP), Campus Assis, Av. Dom Antônio, São Paulo 2100, Brazil

³ Universidade Estadual Paulista Julio de Mesquita Filho, Assis, São Paulo, Brazil

Introduction

The field of assisted reproductive technology (ART) has grown exponentially since the late 1970s. The birth of Louise Brown in June 1978 was a milestone for the *in vitro* fertilization (IVF) technique. This event established the successful path that ART would follow. Since that time, approximately six million babies have been born through ART [1].

Infertility is defined as failure to establish a clinical pregnancy after 12 months of regular, unprotected sexual intercourse. It is estimated that infertility affects as many as 186 million people [2], both males and females, or between 8 and 12% of all couples of reproductive ages worldwide [3].

ARTs include a wide range of different techniques and procedures used to treat sub/infertility or genetic problems that can impede the natural conception of a child. Procedures on the ART spectrum are conducted by manipulating human gametes or embryos *in vitro*. The complexity and difficulty of these techniques vary immensely.

Depending on the reason for using the ART, different procedures can be performed. Some of them are gamete intrafallopian transfer, intracytoplasmic sperm injection (ICSI), IVF, and embryo transfer (ET) [4]. The most common procedures are ICSI, IVF, and ET [4].

In a typical cycle of IVF, failure can occur at several steps, for example the choice of low-quality embryos. Other, indirect factors can influence the outcome of the procedure, such as changes in the physical and psychosocial environments [5]. Mathematical models could be used at certain points in procedures on the ART spectrum, in order to resolve possible problems.

The field of reproductive medicine has been reinventing itself over the years, with continual introduction of new techniques to implement and enhance medical science. Technologies such as the genetic screening test and cryopreservation are outstanding achievements and in certain ways revolutionary for the field of reproductive research. Introduction of the mathematical techniques of artificial intelligence (AI) into the ART could lead to another revolution.

Since the ART techniques involve a set of complex steps and many different variables, this is an ideal and vast field for the application of AI techniques. In recent years, several algorithms have been applied to the classification or prediction of reproductive medical data, applying classical statistical methods such as logistic regression [6, 7], or in other cases using different AI techniques [8, 9].

According to He et al. [10], AI “involves the development of computer algorithms to perform tasks typically associated with human intelligence.” AI is a generic term that can be separated into specific sub-areas, such as artificial neural networks (ANNs), fuzzy logic, genetic algorithms (GAs), deep learning, and machine learning (ML).

AI is one of the key research areas in computer science. With its rapid technological advancement and vast numbers of applications, AI is rapidly becoming pervasive. This increasing use is due mostly to its robust applicability to problems that cannot be easily solved by humans or traditional computing structures [11]. AI tools can process and analyze massive amounts of data within a short time [12].

The history of AI dates back a few decades (Fig. 1). The development of the first artificial neuron by McCulloch and Pitts [13] can be seen as the core source of many of today’s algorithms and a starting point for the field of AI. Since the creation of the term AI in 1956, much has changed. The technique has evolved, and the general public’s perception of AI itself has also changed.

Even before its popularization, AI was already producing outstanding achievements. In 1997, the historic match between Garry Kasparov and IBM’s Deep Blue took place [14], when for the first time a machine outplayed a human in chess. AI has only become widely recognized within the past few years. The growth and development of AI can be seen in some of the most popular technologies that underpin our everyday tasks. Google, Siri, Cortana, and Netflix are some examples of commonly used AI applications.

Research in AI has evolved substantially, due to new hardware that increases the capacity of computer processing; expands research on deep learning (DL) and convolutional neural networks (CNN); and improves the processing of AI models such as artificial neural networks (ANNs), support vector machine (SVM), and other models described below.

The combination of AI and statistical techniques to improve the prediction or classification of medical data is promising. This review traces some of the advances in the combination of AI techniques and ART, and briefly introduces some of the most frequently used algorithms and their functionality.

Methods and design

Information sources and search strategy

A bibliographic search was performed on the following platforms: Google Scholar, PubMed, Capes Periodicals, and Scopus. The search keywords were “assisted reproduction”, “IVF”, “ART”, “embryo quality”, “oocyte quality”, and “semen quality” combined with “artificial intelligence”, “artificial neural networks”, “deep learning”, “convolutional neural networks”, “machine learning”, “Bayesian networks”, and “support vector machine”. Articles published from 2010 until early 2020 were included in the first stage of selection.

Fig. 1 Timeline illustrating key events in AI and ART history

Eligibility criteria

Inclusion criteria

The review focused on studies that applied artificial intelligence techniques to predict or classify reproductive data, such as evaluation of embryo, oocyte or semen quality, and innovations such as predictions of pregnancy or live-birth rates. The review included published observational studies such as cohort study, case-control study, cross-sectional study, case reports, and series. Only journals with a selective editorial policy and tracked by Clarivate, Scopus, or SciELO were considered.

Exclusion criteria

We excluded studies that were performed with nonhuman species; published in languages other than English or Portuguese; and applied linear or logistic regressions, reviews, and editorials and congress abstracts.

Study selection

After the duplicates were removed, two investigators independently assessed the titles and abstracts of all articles. Studies that did not meet the inclusion criteria were

excluded. After this first step, the text of the selected studies was evaluated fully. Disagreements regarding inclusion were resolved by consensus or with the involvement of a third author.

Data extraction

The data extracted from all selected articles included author(s), AI technique used, type of data used in AI, number of samples, selected features used, objective in the application of AI, and results obtained. Articles that lacked one or more of these information categories were excluded.

Results

Literature searches and study selection

The searches resulted in 124 articles; after duplicates were removed and the exclusion criteria applied, 76 abstracts were screened. A full-text review was performed for 46 articles. In this stage, reviews ($n = 10$), articles that did not provide results ($n = 3$), and studies that did not specify the AI used ($n = 1$) were excluded. After all selection steps, 32 articles met the inclusion criteria and were included in this review, as illustrated in the flowchart (Fig. 2).

Fig. 2 Flowchart of the literature searches and study selection

Artificial neural networks

The neurons in a human brain form an intricately interconnected network. They are highly specialized cells in the reception and conduction of electrical impulses, receiving and transmitting stimuli, and being able to transmit information (as electric pulses) to nearby and connected neurons. The learning process occurs by activating these cells, through the propagation of action potentials, where there is a difference in membrane electrical potential. This is a poorly understood and complex biological process. However, the ANNs attempt to simplify and mimic this brain behavior. This makes these ANNs' neurons (called nodes) more likely to produce a desired outcome given a specific input [15].

The ANN is connected hierarchically, which means that the output of some nodes can be the input to others. This network can be represented through layers of nodes. The processing ability of the network is stored in the interunit connection weights, present in the nodes, that enable the learning process. Each node applies an activation function to the sum of the received weighted inputs to produce an output (Fig. 3). Depending on the resulting value, the node can be activated or not [15].

The layers of the ANN constitute the architecture of the model. The way that those nodes are connected, and in which layers, depends on the type of ANN used. The main type of ANN is the multilayer perceptron (MLP).

The simplest architecture of the MLP consists of three layers, with one hidden layer. Importantly, all the nodes in

all the layers have the same general purpose, to sum the weights of the neural network and pass through an activation function. Although all the nodes have the same purpose, the existence of the hidden layers and the way that they are connected profoundly influence the performance of the model.

As seen in Fig. 4, the main difference between the perceptron and the MLP is in the presence of the hidden layer. This extra layer substantially increases the performance of the MLP. The MLP is capable of solving linear and nonlinear separable problems, whereas the perceptron is capable of solving only linear problems.

In the field of assisted reproduction, many problems can be solved by using the ANN. This technique has a wide range of applications because it can be used to solve problems involving classification, prediction, or selection of samples. For instance, in the ART, the classification and selection of oocytes,

Fig. 3 Representation of an ANN node

Fig. 4 Representation of the architectures of three different ANN models

sperm, and embryos are routinely conducted by embryologists and can be facilitated with AI-based tools.

Table 1 lists the main studies that have used the MLP architecture of the ANN to classify different types of ART data. Most of the articles used a similar three-layer MLP architecture for the network, where only the number of nodes in each layer varied. There is no set of rule for choosing the number of nodes, since each problem has its own specificity and complexity.

The selected papers illustrate the diversity of the types of data applied to the multilayer perceptron networks. Three of these [18–20] attempted to determine semen quality through the use of information on the lifestyle and environment of the patients. All of them aimed to correlate a man's lifestyle with abnormalities in the semen, through variables such as age, smoking and drinking habits, season of the year, and allergies. Girela et al. [18], for example, had 34 variables, but for AI, more variables do not equal better results. Irrelevant or redundant data can have a negative impact on the learning and processing speed of the model.

Feature selection is an important data pre-processing step, which may profoundly impact the performance of the AI model. This process consists of determining the relevant data and removing the irrelevant, redundant, and noisy data that offer no useful information [25]. Sahoo and Kumar [20] originally had nine variables to use as inputs in the classifiers, but after feature selection, this number was reduced to seven. The authors concluded that childhood diseases and high fevers are less relevant and contribute less to the prediction of semen quality.

Durairaj and Thamilselvan [16] and Hassan et al. [22] targeted their approach not only on the embryo itself but also on the other factors involved in the success of IVF treatment. They used variables that focus more on women's physiology, and also used variables for men, such as sperm concentration, and variables about the ART cycle, such as the number of

oocytes retrieved and the embryo transfer day. Hassan et al. [22] applied four other classifiers besides MLP. Importantly, after the feature selection, each classifier had a different set of variables, because the variables have different effects on each classifier. However, important features such as the age of the patient appeared in all five sets of variables.

With the rise of time-lapse techniques in ART laboratories, a new perspective to the development of the embryo was revealed and a new way to analyze its competence has also emerged. Time-lapse monitoring generates a huge amount of data that can and has been analyzed by the AI techniques. Milewski et al. [17] used six morphokinetic parameters: t2, t3, t4, t5, with absolute values, and two relationships, cc2 (t3-t2) and s2 (t4-t3), for the prediction of embryo implantation (Table 2). The MLP can perceive patterns in the data and can make predictions based only on the kinetic information on the embryo, a task that could not be performed by a human.

In two of the articles, the MLP was combined with another algorithm to optimize the process. Milewski et al. [17] used the principal component analysis, which provides information about any common structure that might exist in a dataset. El-Shafeiy et al. [19] used the Sperm Whale algorithm, which consists of an optimization algorithm based on the sperm whale's lifestyle. Both algorithms have the purpose of increasing the ability of the MLP to learn from the data and improve accuracy [19, 26].

Miyagiet et al. [21] applied six different classifiers to the task of predicting a live birth; the logistic regression gave the best result. Usually, the MLP would give a better result than the logistic regression because it can work with nonlinear functions and complex problems, being a more sophisticated type of AI and considered a universal approximator.

The universal approximation theorem states that an MLP can compute any kind of function, and that even with only a single hidden layer, the power of accuracy can be extremely high, terming these networks “universal approximators” [15,

Table 1 Studies that applied MLP to reproductive data

Reference	Type of data	Number of samples	Selected features	Desired outcome	Reported performance
Durairaj and Thamilselvan [16]	IVF cycle, sperm, and woman	Dataset for 250 patients	Variables regarding the woman, man, and cycle	Success rate of IVF	75% accuracy
Milewski et al. [17]	Embryo	610 embryos fertilized by intracytoplasmic sperm injection from 514 cycles	Morphokinetic data from the embryos	Embryo implantation	0.71 AUC in validation test
Girela et al. [18]	Semen samples and semen	Dataset for 123 patients	34 variables regarding the environment and lifestyle of the volunteers	Quality of the semen	69% accuracy
El-Shafeiy et al. [19]	Semen	Dataset for 100 patients	9 variables regarding lifestyle of the patients	Quality of the semen	99.96% accuracy using SWA
Sahoo and Kumar [20]	Men	Dataset for 100 patients	Variables regarding environment and lifestyle of the volunteers	Quality of the semen	92% accuracy
Miyagi et al. [21]	Embryo image	160 images of blastocysts	Variables regarding blastocyst image	Live birth	65% accuracy
Hassan et al. [22]	Cycle, sperm, and woman	Dataset for 1048 IVF patients	Variables regarding the woman, man, and cycle	Success rate of IVF	97.7% accuracy
Manna et al. [23]	Oocyte and embryo images	269 oocyte images and 269 embryo images	Variables regarding oocyte and embryo image	Oocyte/embryo quality scoring	Approximately 0.8 AUC
Candemir [24]	Semen samples	Data on 100 healthy volunteers	Environment and daily life habits of the volunteers	Quality of the semen	90% accuracy

27]. According to Hornik et al. [27], lack of success in application of the model arises from inadequate learning, an insufficient number of nodes, or the lack of a deterministic relationship between the inputs and the outputs. It is presently understood that several factors can cause deficient learning, among them splitting data inappropriately, hiding variables, and mistaking the objective [28]. These factors could be solved with greater effort to understand the operation of algorithms and the appropriate way to utilize them [28].

As seen in Fig. 4, the radial basis neural network (RBNN) has a similar architecture to a three-layer MLP, but the main difference lies in the activation function used. In the RBNN, only the radial basis function is used for the activation.

Although they used different types of AI models, Girela et al. [18], El-Shafeiy et al. [19], and Candemir [24] applied similar types of data to the task of classifying semen data. All of them used the life habits of male patients and achieved very

positive results (Table 1). One of the possible reasons for this high accuracy may be the type of data used. The possible values that each variable can assume are not very diverse and are more direct, making it much easier for the classifier to group the data and make a classification.

Bayesian networks

A Bayesian network (BN) is a probabilistic graphical model. This AI technique, in contrast to the other methods described in this review, is not merely a method employed in AI but also a classical and well-known statistical model [29]. BNs aim to represent a given set of variables and their conditional relationship of dependence through a directed acyclic graph. The architecture of the model consists of nodes, in which a node is associated with a random variable and a conditional probability from its parent node, the variables can be continuous or discrete, and one node can have more than one variable and effect on other nodes (Fig. 5).

The BN works well with incomplete data, in a way that allows the machine to make inferences and predictions, in order to make decisions regarding the probability of occurrence of different events such as the possibility of a pregnancy and live birth in the ART. Because the BN is graphically represented, it is easier to interpret and more intuitive. This feature of the BN can facilitate the interaction between the physician and the AI. Some examples of the application of the BN models to predict reproductive data in ART are summarized in Table 3.

Table 2 Morphokinetic parameters used by Milewski et al. [17]

Initials	Meaning
t2	Time from insemination to second division
t3	Time from insemination to third division
t4	Time from insemination to fourth division
t5	Time from insemination to fifth division
cc2	t3-t2: second round of cleavage
s2	t4-t3: second synchronization parameter

Fig. 5 Representation of two different BN architectures. **a** Naïve Bayes architecture. **b** Tree-augmented naïve Bayes architecture

A recurrent problem when working with databases is missing data. Gianaroli et al. [29], Corani et al. [32], Hernández-González et al. [30], and Morales et al. [31] described the problem of partially observable variables, when the fate of a specific embryo is unknown. The variable is considered partially observable because the number of implanted embryos is smaller than the number of transferred embryos, which leads to the missing-data problem.

Each of the authors followed a different strategy to circumvent the problem. Gianaroli et al. [29] and Corani et al. [32] used the averaging approach, while Hernández-González et al. [30] used the learning from label proportions. Both techniques allow the use of the unlabeled embryos and avoid data loss. Morales et al. [33], facing the same problem of partially observable variables, resolved this problem with an approach to assume which of the embryos had been implanted, using the criterion of the highest score from the transferred batch. When working with a supervised type of learning, this means that all the samples must have a label. It is important to not have this kind of missing data, because the classifiers would not be able to associate the sample with a specific type of outcome and consequently would not be able to learn from that sample.

Several different architectures and variations of the BN were presented in the selected papers. Among the most often used were the naïve Bayes and tree-augmented naïve Bayes

(TAN); some of the other architectures used were the selective naïve Bayes, semi naïve Bayes, and K-dependence Bayesian classifiers (KDB). There is an extensive range of the BN models, all of which feature the relationship of dependence and independence between variables, but they differ and vary in their complexity and efficiency.

Morales et al. [31] tested seven different models (naïve Bayes, selective naïve Bayes, semi naïve Bayes, TAN, KDB, TAN-wrapper, and KDB-wrapper). The best results for accuracy were achieved by the TAN and KDB architectures. These results can be easily explained by the difference in the complexity of the algorithms. Taking the naïve Bayes and TAN models as an example, the main difference between them is the assumption of the relationship of conditional independence of the predictor variables in the given class.

In the naïve Bayes, all the predictor variables of a class are independent from each other, while in the TAN, the dependence between predictor variables is taken into consideration, thus making the algorithm more suitable for complex problems. However, Uyar et al. [34], after testing six different classifiers including the MLP and radial basis function network, found that the model with the best accuracy was the naïve Bayes.

The way that the BN is molded through the directed acyclic graph makes it possible to select specific nodes and their

Table 3 Studies that used Bayesian network models to analyze reproductive data

Reference	Type of data	Number of samples	Selected features	Desired outcome	Reported performance
Hernández-González et al. [30]	Embryo	330 consecutive cycles and associated embryo dataset	Variables regarding the cycle and the embryo	Embryo implantation	81% accuracy
Morales et al. [31]	Embryo images	249 embryo images	Embryos were catalogued according to Mill's score	Embryo implantation	0.8918 AUC
Gianaroli et al. [29]	Embryo	Embryos from 338 cycles	Variables regarding the woman, man, and embryo	Pregnancy	81.5% accuracy and 0.72 AUC
Corani et al. [32]	Embryo	Embryos from 338 cycles	Variables regarding the woman, cycle, and embryo	Pregnancy	0.732 AUC
Morales et al. [33]	Embryo	189 embryos from 63 cycles	Variables regarding the woman, man, cycle, and embryo	Embryo implantation	68.25% accuracy
Uyar et al. [34]	Cycle	Data for 2453 embryos	Variables regarding the woman, cycle, and embryo	Embryo implantation	80.4% accuracy

parent variables; in other words, it is possible to determine the specific variable that affects the outcome of the node, such as the age of the woman (which could affect the receptivity of the uterus), or the embryo score, which affects the embryo viability.

Support vector machine

Support vector machines—SVMs—are based on statistical learning theory and have a great advantage in solving small-sample, nonlinear, and high-dimensional classification problems. The algorithm can be used in both classification and regression problems [35]. The mechanism of SVMs is illustrated in Fig. 6; in this model, the data can be mapped through the hyperplane.

The model implements a nonlinear mapping from the input data to a characteristic space. The hyperplane is then built to separate data into various classes [35]. This being so, the best hyperspace is the one containing the maximum separation space between the classes [36]. Simply stated, the SVM can be used mainly for the classification and separation of samples into two or more categories. Table 4 summarizes the articles that used this AI technique in assisted reproduction.

When manipulating images, one of the most important steps in modeling the project is pre-processing and treatment of the images. This step can include image segmentation, rotation, removal of the background, and any other manipulation used. Poor execution of this step may lead to low accuracy of the ML model, because of the inadequately generated database. The next step is extraction of features from the images; these features are the information that will compose the database from which the AI algorithm will learn.

Santos Filho et al. [37], Wang et al. [39], and Xu et al. [40] used images of embryos in an SVM model, with the purpose of selecting the embryos with the highest potential for implantation. Each of these teams used a different approach to extract

features from the images. Wang et al. [39] and Xu et al. [40] applied the local binary pattern (LBP) algorithm, a feature-extraction method that describes the texture and shape of an image; the texture is represented as a regular repetition of an element or pattern on the surface of an image [41]. The algorithm operates by dividing the image into small segments from which the features are extracted [42].

Wang et al. [39] also applied the gray-level co-occurrence matrix (GLCM), another texture-extractor algorithm. The algorithm works by analyzing the distribution of gray values through the image, because the spatial distribution of gray values is one of the defining qualities of texture [41].

Santos Filho et al. [37] took a different route to analyze images of human embryos. Image segmentation was analyzed in three parts, the zona pellucida, trophectoderm, and inner cell mass. Each image segment was analyzed by different standards: the “thickness” and regularity of the trophectoderm was calculated and the texture of the inner cell mass was extracted by two algorithms, a texture spectrum that measures the symmetry of a region [10, 37, 39] and a gray-level run length-based algorithm that measures the frequency of occurrence of long runs of a given gray level in a region of interest [37, 43]. Santos Filho et al. [37] also opted to train three different classifiers, one to grade the development quality, one to grade the inner cell mass quality, and one to grade the trophectoderm quality. The desired output was the grading system of Gardner and Schoolcraft [38].

Uyar et al. [34] compared six different classifiers with the same database (naïve Bayes, K-nearest neighbor, decision trees, SVM, MLP, and RBFNN) to predict the outcome of IVF implantation. The decision tree provided the poorest AUC result. This can be explained by the simplicity of the model, which does not work well with complex problems. The naïve Bayes gave the best result, and the SVM was the fourth-best AUC.

Convolutional neural network

Deep learning is a modern extension of the classical neural network technique, capable of building neural networks with a larger number of layers. That is possible due to the rapid development of modern computing. Deep learning can explore more complex nonlinear patterns in the data, which explains why deep learning is mostly used in imaging analysis, considering that images are naturally complex and require extensive computational processing [44].

Deep learning is an umbrella term that includes several different types of algorithms, such as recurrent neural network, deep belief networks, and convolutional neural network (CNN) [44]. The name “convolutional neural network” indicates that the network employs a mathematical operation called convolution, which is a specialized kind of linear operation. Convolutional networks are simply neural networks that

Fig. 6 Representation of the architecture of the SVM model, illustrating the data mapping through the hyperplane

Table 4 Reproductive data applied to an SVM classification model

Reference	Type of data	Number of samples	Selected features	Desired outcome	Reported performance
Santos Filho et al. [37]	Embryo images	73 embryo images	Segmentation of the blastocyst image	Blastocyst grading according to the system of Gardner and Schoolcraft [38]	53–92% accuracy in expansion, inner cell mass (ICM), and trophectoderm (TE)
Wang et al. [39]	Embryo images	206 embryo images	Local binary pattern and gray-level co-occurrence matrix	Embryo implantation	84.72% in cross-validation and 77.67% in test set
Uyar et al. [34]	Cycle data	Data for 2453 embryos	Variables regarding the woman, cycle, and the embryo	Embryo implantation	0.657 AUC
Xu et al. [40]	Embryo images	185 embryo images	Textures of embryo image	Embryo implantation	Approximately 65% accuracy

use convolution in place of general matrix multiplication in at least one of their layers [45] (Fig. 7).

The CNNs resemble MLPs in many ways; they have nodes that learn based on weights and receive several different inputs to sum and pass through an activation function to produce an output. In image analysis, convolution is a mathematical operator that takes an image and a filter as input and produces a filtered output [46]. Because of this potential, many studies have used the CNN in different fields. Table 5 lists the articles found that applied the CNN to some sort of image classification.

The convolution layer is the core of the CNN, where the main portion of the network computation occurs. An image consists of a matrix of elements; the core of the CNN is a smaller matrix that has a specific task such as detecting lines or edges. The CNNs can be used for tasks such as object detection and semantic segmentation. The CNNs are used primarily for images, such as in the process of classification and clustering. For this reason, all the selected papers dealt with the application of the CNNs for some type of image classification.

Khan et al. [47], Rad et al. [54], and Patil et al. [46] used the model to count cells in microscopic images but with different purposes. Khan et al. [47] applied the CNN to time-lapse images to count the number of cells in developing embryos up to the five-cell stage. Rad et al. [54] used the CNN to count the number of blastomere cells and identify their centromeres, while Patil et al. [46] used a model to automate the process of grading embryos from day 1 to day 5.

Matusevičius et al. [49] combined the CNN algorithm with the SVM for the task of detecting cells from embryo images. The SVM algorithm was used to detect cell regions, which were later used to train the CNN model. Dirvanauskas et al. [51] went further and wanted to determine the stage of development of the embryo. For this purpose, they used a modified version of the CNN called AlexNet to determine the stage of development of the embryo, achieving more than 97% accuracy.

The process of counting cells has been used to classify embryos regarding their day of development. All the authors agreed that the task of analyzing images of embryos throughout their development is time-consuming and inevitably affected by the subjectivity of the human analyst, intra- and inter-individual unpredictability, and human error [46]. This demonstrates the potential impact and importance of an automated system that would eliminate these problems.

In a different approach, Chen et al. [48], Cao et al. [52], VerMilyea et al. [57], and Kragh et al. [58] used embryo images to train AIs for a classification, based on different grading systems. Chen et al. [48] and Kragh et al. [58] used the grading system of Gardner and Schoolcraft [38]. Kragh et al. [58] combined two different AI techniques, CNN and recurrent neural networks, to grade the inner cell mass and trophectoderm.

Chen et al. [48], using a larger dataset, applied the CNN for a three-part classification of blastocyst development, inner cell mass, and trophectoderm. They achieved an overall better performance when comparing the two studies [48, 58]. VerMilyea et al. [57] and Cao et al. [52] had the purpose of predicting embryo viability. Despite using the smaller database from Cao et al. [52], they achieved a better result when applying the AI technique. These differences make it clear that the size of the dataset is not always a meaningful factor in obtaining the best accuracy.

Khosravi et al. [53] and Bormann et al. [56] aimed to determine the quality of the embryo. Both utilized a deep neural network, a type of neural network with several hidden layers. Bormann et al. [56] used a grading system with five grades (from poor to excellent), while Khosravi et al. [53] used only three grades (poor, fair, and good). Perhaps because of the fewer categories (grades) and more data used, Khosravi et al. [53] obtained a better result.

Jiang et al. [44] defined the CNN as a type of feed-forward artificial neural network consisting of nodes that have learnable weights and biases, very similar to the ordinary MLP networks. Despite these similarities, the CNN is part of a

Fig. 7 Representation of a CNN structure

more-robust type of AI algorithm. While the typical MLP has three layers, the CNN architecture used by Matusevičius et al. [49], for example, had 15 layers. The architecture of the CNN is more suitable than an MLP for classifying images. However, this type of AI has a higher computational cost than MLP.

Miyagi et al. [50] and Tran et al. [55] had the most daring goals in this section. Miyagi et al. [50] previously made predictions of live birth using embryo images with an MLP model but achieved an increase in accuracy only by using a considerably larger set of images and a different model. Although this increase was not significant, predicting a live birth is extremely difficult and is the gold standard pursued by all research in this field. Tran et al. [55] wanted to predict the probability of a pregnancy with fetal heartbeat. This endpoint (fetal heartbeat) is slightly easier to predict than the live-birth endpoint, but is more difficult than merely predicting the embryo quality. Both endpoints (fetal heartbeat and live birth) have several intervening variables that play a role in the resulting pregnancy, and that do not depend solely on the morphology or quality of the embryo.

Clinical applications of AI techniques

Based on the AI techniques described in the above-cited articles, the following tables present a synthesis of the types of data used, the IVF technique, the outcome prediction, the AI methods applied, and the performances of these methods, as well as examples of their strengths and weaknesses, both from the point of view of the database used and the results obtained

from AI, and from their possible clinical applications in IVF laboratories.

Focusing on embryo selection, the use of AI techniques suggests an improvement in the results—or at least, an automation of the process. In all cases, the main objective is to develop systems that are able to automatically support the decision of which embryo to transfer. In addition, the systems can improve the success rates of infertility treatments in the clinics, helping to make correct decisions in complex cases and reducing the incidence of twin pregnancies by selecting a single embryo for transfer.

In terms of applicability, the use of graphical representations such as the Bayesian classifiers can facilitate the process of understanding the probabilities of the classification processes, aiding in interpretation by clinicians [29–33]. The automated systems can also be used to train new embryologists. Other benefits of the automation process are the shortening of the time required for embryologists' evaluations, providing a rapid diagnosis that can be performed remotely, without the need to be physically present in the laboratory [33]. The variables from cycle data (alone or combined with other data) can be a useful source of information (Table 6).

Despite allowing noninvasive and accurate embryonic analyses, some studies should be interpreted with caution with respect to certain aspects. Some of them evaluated the data retrospectively and only a few were able to compare the results obtained with the rates of clinical results for implantation or pregnancy [29, 34, 40]. In this review, we considered a database as “small” when it had fewer than 100 samples, a “relatively small” database as having 100 to 500 samples, and

Table 5 Studies that used a CNN to analyze reproductive data

Reference	Type of data	Number of samples	Selected features	Desired outcome	Reported performance
Khan et al. [47]	Time-lapse	20 embryo time-lapse sequences	Variables regarding the segmentation of embryo image	Embryo segmentation	84.85% accuracy
Patil et al. [46]	Embryo image	500 samples of embryo images	Variables regarding embryo image	Embryo classification	85.7% accuracy
Chen et al. [48]	Blastocyst image	171,239 images from 16,201 embryos	Variables regarding blastocyst image	Blastocyst grading according to the system of Gardner and Schoolcraft [38]	75.36% accuracy in all three categories
Matusevičius et al. [49]	Embryo image	300 embryo images	Variables regarding embryo image	Embryo detection	11.92% mean square error
Miyagi et al. [50]	Blastocyst image	5691 blastocysts	Variables regarding blastocyst image and age of patient	Live birth	72% accuracy
Dirvanauskas et al. [51]	Time-lapse	7002 embryo images	Variables regarding embryo image	Embryo development stage	97.62% accuracy
Cao et al. [52]	Embryo image	1310 images from 344 embryos	Variables regarding embryo image	Embryo viability	78.14% accuracy
Khosravi et al. [53]	Embryo image	10,148 embryos	Variables regarding embryo image	Embryo quality	97.53% accuracy in the blind test set
Rad et al. [54]	Embryo image	224 embryos	Variables regarding embryo image	Blastomere cell counting and centroid localization	88.2% for embryos of 1–5 cells
Tran et al. [55]	Time-lapse	10,638 embryos	Variables regarding embryo image	Pregnancy predictor	0.93 AUC
Bormann et al. [56]	Embryo image	748 embryos	Variables regarding embryo image	Embryo quality	83.92% consistency
VerMilyea et al. [57]	Embryo image	8886 embryos	Variables regarding embryo image	Embryo viability	> 63% accuracy in the blind test set
Kragh et al. [58]	Embryo image	8864 embryos	Variables regarding embryo image	TE and ICM grading according to the system of Gardner and Schoolcraft [38]	65.2% and 69.6% accuracy for ICM and TE grading, respectively

a “large” database as containing more than 500 samples. Another important issue is that some studies only considered embryonic data and did not compute clinical or patient data, which may be important for increasing the accuracy, for example [29, 30, 34] (Table 7).

Increasingly, AI methods have been used in the ART in order to improve the reproductive outcomes. Oocyte and embryo selections can be considered the starting point for improving the success rate in IVF clinics: Manna et al. [23] evaluated the [quality of oocytes](#) and the respective embryos through AI, and although they reached a slightly weak AUC, their proposed system could help medical professionals to select the best-quality oocytes and embryos. [One of the limitations](#), cited [by the authors](#), refers to the use of static images obtained through optical microscopy, which may prevent a comprehensive prediction of oocyte and embryo qualities.

Several studies used static images of embryos for AI analysis, e.g., Morales et al. [31]; Santos Filho et al. [37]; Wang et al. [39]; Xu et al. [40]; Patil et al. [46]; Dirvanauskas et al. [51]; and VerMilyea et al. [57], with promising results.

However, the use of a time-lapse system could provide a larger dataset of images from a single embryo [53, 55, 56, 58], in addition to the dynamics of embryo development, or morphokinetic data [17]. Although it has shown promising results, the use of a time-lapse system must be carefully considered. This technology is not widely used in clinics worldwide, and its cost can be remarkably high. Moreover, the proper definition of the morphokinetic markers may vary with the clinician and still requires an automated and unbiased process. Table 8 summarizes the publications that used time-lapse data.

Sperm selection is also a decisive criterion to improve the prediction of male fertility and ART outcomes. Most of the publications involving semen quality attained high accuracy [19, 20, 24], but they also showed limitations such as small datasets and unbalanced data (Table 9). These factors may impact the final accuracy of the models; it is important to have a considerable quantity of data to train and use unbiased data, so that the AI system correctly learns all types of clinically meaningful categories. Even so, the models have benefits that can facilitate sperm evaluation and may be useful to diagnose

Table 6 Comparison among different AI methods for ART reproductive outcomes, based on IVF cycle data. *AUC*, area under the curve; *ROC*, receiver operating characteristic curve

Reference	Type of data	Technique	Outcome prediction	Type of AI	Performance	Strengths	Weaknesses
Durairaj and Thamilselvan [16]	Cycle, woman, and man	IVF/ICSI	Success rate of IVF	ANN	75% accuracy	Included variables from both members of the couple could help physicians in counseling treatments	Did not include variables related to embryo quality, relatively low accuracy
Hassan et al. [22]	Cycle, semen, and woman	IVF/ICSI	Success rate of IVF	ANN	97.7% accuracy	High accuracy and large dataset could help to guide medical professionals in assessing pregnancy and counseling treatments	Did not include variables related to embryo quality
Hernández-González et al. [30]	Cycle and embryo	IVF/ICSI	Embryo implantation	BN	81% accuracy	Inclusion of the cycle features increased the classification performance	Limited predictive capability of the variables; classification was difficult for medium-quality blastocysts. Relatively small sample; retrospective study
Gianaroli et al. [29]	Cycle	IVF/ICSI	Pregnancy	BN	81.5% accuracy and 0.72 AUC	The model discriminated pregnancy, single pregnancy, and twin pregnancy.	
Corani et al. [32]	Cycle and embryo	IVF/ICSI	Pregnancy	BN	0.732 AUC	The model was equally or more predictive than well-recognized classification algorithms; model is easy to interpret.	Relatively small sample (non-synthetic); static observation of embryos
Morales et al. [33]	Cycle	IVF/ICSI	Embryo implantation	BN	68.25% accuracy	More accurate selection than present evaluation procedures; the graphical structure of BN facilitated interpretability and understanding by clinicians.	Retrospective study did not apply ROC curves—non-embryo-centered analysis—and did not take into account single embryo transfer
Uyar et al. [34]	Cycle	ICSI	Embryo implantation	BN, SVM	80.4% accuracy	Large database: acceptable accuracy level greater than that provided by evaluation of embryologists could aid embryologists in making the right decisions in such complex cases	Retrospective study included only day 2 and day 3 cleavage stage-embryos and did not guarantee the selection of embryos because it did not provide confidence level

Table 7 Comparison among different AI methods for ART reproductive outcomes, based on oocyte and/or embryo images. *AUC*, area under the curve; *ICM*, inner cell mass; *TE*, trophectoderm; *CE*, conventional evaluation

Reference	Type of data	Technique	Outcome prediction	Type of AI	Performance	Strengths	Weaknesses
Miyagi et al. [21]	Embryo image	IVF	Live birth	ANN	65% accuracy	Balanced data could help to select best-quality embryo	Small database: low accuracy
Manna et al. [23]	Oocyte and embryo images	ICSI	Oocyte and embryo qualities	ANN	0.8 AUC	Could help physicians and biologists to select the best qualities of oocytes and embryos	Small database: result reached by AUC slightly weak
Morales et al. [31]	Embryo image	ICSI	Embryo implantation	BN	0.8918 AUC	Graphical structure of BN could facilitate interpretability and understanding by clinicians	Small sample: only two algorithms achieved better performance compared with the seven presented (need to improve accuracy of classification)
Santos Filho et al. [37]	Embryo image	IVF	Blastocyst grading according to system of Gardner and Schoolcraft [38]	SVM	53–92% accuracy in expansion, ICM and TE	More accurate and non-subjective approach for embryos within a single grade; subtle quantification of individual morphological features	Small database: more complete and robust classifiers are necessary; poor blastocyst image quality (ICM and TE with weak edges)
Wang et al. [39]	Embryo image	IVF	Embryo implantation	SVM	84.72% in cross-validation and 77.67% in test set	The evaluation model established by the texture feature had a good effect on the prediction	Small database; predictive value of model could not be predicted with certainty.
Xu et al. [40]	Embryo image	-	Embryo implantation	SVM	Approximately 65% accuracy	Extraction of the texture feature allowed generation of two types of linearly separable images	Small database evaluated only embryos on day 3 of development.
Patil et al. [46]	Embryo image	IVF	Embryo classification	CNN	85.7% accuracy	Evaluated embryo images during first 5 days of development	Relatively small sample
Chen et al. [48]	Embryo image	IVF	Blastocyst grading according to system of Gardner and Schoolcraft [38]	CNN	75.36% accuracy in all three categories	Large database	Retrospective study did not include cycle or patient data and did not assess prediction of pregnancy/implantation.
Matusevicius et al. [49]	Embryo image	-	Embryo detection	CNN	11.92% mean square error	Error rate of the model was less than the error rate of embryologists' experience; the model worked better for predicting cell position	Relatively small sample
Miyagi et al. [50]	Embryo image	IVF	Live birth	CNN	72% accuracy	Large dataset; CNN was better than CE regarding sensitivity	Limited to blastocyst stage; no significant difference between CNN and CE regarding accuracy and AUC.
Cao et al. [52]	Embryo image	IVF	Embryo viability	CNN	78.14% accuracy	Improved medical image processing could facilitate embryo selection.	Use of static images: relatively small dataset
Rad et al. [54]	Embryo image	IVF	Blastomere cell counting and centroid location	CNN	88.2% accuracy	Automated method for counting blastomeres could improve prediction of embryo's implantation potential.	Small dataset: limited to cell counting
VerMilyea et al. [57]	Embryo image	IVF/ICSI	Embryo viability, measured by clinical pregnancy outcome	CNN	64.3% accuracy	Large dataset improved predictive value for evaluation of embryo viability compared with embryologists	Use of static images of blastocysts: low accuracy

Table 8 Comparison among different AI methods for ART reproductive outcomes, based on time-lapse data. *AUC*, area under the curve; *ICM*, inner cell mass; *TE*, trophoctoderm

Reference	Technique	Outcome prediction	Type of AI	Performance	Strengths	Weaknesses
Milewski et al. [17]	ICSI	Implantation	ANN	0.71 AUC	Predictive value of morphokinetic data on embryo implantation	Included only morphokinetic data; weak result reached by AUC
Khan et al. [47]	IVF	Embryo segmentation	CNN	84.85% accuracy	Method successfully segmented complex and irregular embryo structures; the segmentation approach made it possible to overcome noise and artifacts from embryo images in the dark field.	Small database; limited to manual segmentation
Dirvanauskas et al. [51]	IVF	Embryo development stage	CNN	97.62% accuracy	High accuracy; could help to improve embryo selection	Limited to prediction of embryo stage
Khosravi et al. [53]	IVF	Embryo quality	CNN	97.53% accuracy	Large dataset; high accuracy	Limited to scoring embryos. The accuracy was measured based on embryologists' internal scoring system
Tran et al. [55]	IVF	Pregnancy	CNN	0.93 AUC	Large dataset	Limited to blastocyst stage
Bormann et al. [56]	IVF/ICSI	Embryo scoring and decisions for biopsy and cryopreservation	CNN	84% consistency	Large dataset; the proposed method for consistency outperformed the embryologists.	The system evaluated was measured by consistency; efficacy still needs to be tested in clinical settings through prospective studies.
Kragh et al. [58]	IVF/ICSI	Embryo quality	CNN	0.66 AUC	Large database; ICM and TE assessments by the proposed method were slightly better than those by the embryologists.	Focused on assessment of ICM and TE qualities; blastocyst morphology prediction by the proposed method was similar to that performed by embryologists.

Table 9 Comparison among different AI methods for ART reproductive outcomes, for the prediction of sperm quality

Reference	Type of AI	Performance	Strengths	Weaknesses
Girela et al. [18]	ANN	69% accuracy	Could be useful for early diagnoses of seminal disorders; could help to select candidate semen donors	Little available data; unbalanced data; low accuracy
El-Shafeyi et al. [19]	ANN	99.96% accuracy	High accuracy; could be a powerful tool to predict male fertility	Small database; unbalanced data
Sahoo and Kumar [20]	ANN	92% accuracy	High accuracy	Small database; variables limited to the environment and lifestyle of the men
Candemir [24]	ANN	90% accuracy	High accuracy; could be used as a pre-test before applying laboratory tests	Limited to daily life habits and environment of the men

seminal disorders or to select semen donors. Besides, a model could be easily integrated into the workflow, at the time of the andrologist's analysis, for example, and more information about the male patients, such as their lifestyle and environment, could be included through questionnaires provided by the laboratories.

Consideration of these articles reveals the need for more research and standardization, as previously stated by Ratna et al. [59]. Larger databases should be used in order to improve the robustness of the model's performance. For high-performance studies, it is still necessary to test them prospectively, applying the classification systems in clinical practice, in order to assess the future success rates of infertility treatments and to validate them in actual clinical practice.

In general, we found that there is a need for research with broader and unbiased databases and applied in several clinics at the same time, in order to validate the models in different clinical contexts (i.e., robustness and reproducible validation). Notably, no article cited in this review reached the “Holy Grail” of this research field: that is, a proof of concept, based on a prospective selection and entirely based on the AI tool. This would be essential to validate the models in the clinical context and open the way to the routine use of AI in ART clinics.

Discussion

With the rapid advances in the field of AI and ML, the existence of these techniques cannot be overlooked. It is important to comprehend the wide range of possibilities for applying the different techniques and to have a basic knowledge of the process of prediction and classification through these algorithms. For the ART, there is an enormous amount of different types of data that can be used for the AI algorithms, as well as a range of possible combinations of algorithms and architectures to best fit to a problem.

For solving a particular problem, there is no single correct AI method or one that fits perfectly in all cases. Some key points to consider are the type and amount of available data, whether labels are available, and the type of output desired.

All this can be decided only through careful analysis of the dataset; even so, more than one algorithm may be appropriate to solve the problem.

For example, in comparing a BN to an MLP, although the Bayesian approach is usually simpler than the neural network, it still has its advantages. The visual representation of the graph of the Bayesian network itself has meaning, and the architecture of the model provides important information about the variables' relationships and their conditional dependence, whereas the MLP does not allow such a direct and explicit interpretation. However, when a dataset is large and composed of images, for example, the architecture of a CNN may be preferred due to its high classification power when manipulating this type of data.

Another core step when dealing with AI is the correct modeling of the problem, which will help to improve the technique. For correct modeling, it is important to carefully choose the AI technique that best fits the problem. This technique must also be used correctly to achieve the best outcome. Most of the time, the lack of a good result can be traced to a poorly constructed database. The model will learn based on the samples that are provided; if the data are biased, with noisy and irrelevant information, or a class of samples is under-represented, for example, the algorithm may learn incorrectly, leading to low accuracy when new data are entered.

There are many ways to combine the ART with AI, and this contribution has discussed some of these possibilities and their applications. As previously commented, objective assistance during the selection of the best embryo to be transferred may be one of the main applications of AI. In this sense, after evaluating the performance and methodological quality of the proposed models for the prediction of in vitro fertilization's issues, we suggest the need for prospective validation studies of those algorithms generated by AI techniques. For high-performance models, the algorithms require additional external validation to assess their applicability and robustness (i.e., the same accuracy in different conditions apart from those that they were generated). Moreover, prediction algorithms can be integrated into a user-friendly interface to facilitate their handling by clinicians to assist the decisions [59].

Other possibilities can also be explored, especially with the current and fast-evolving technology. Analysis of the current research indicates that in a short time, artificial intelligence will be used much more often in laboratories and IVF clinics than we could imagine. In a sort of fictional speculation, the AI tools could reach the threshold of application to the identification of euploid embryos, which the best embryo to fresh or frozen transfer, the embryos that single or coupled would establish the high probability of singletons etc. Thus, it is essential that medical professionals are increasingly familiar with artificial intelligence. Certainly, it will be integrated soon into clinical routine. Possibly, it will be a revolutionary new milestone in the assisted reproduction service.

Acknowledgements Grant # 2017 / 19323-5, 2018 / 24252-2 and 2018 / 19371-2 São Paulo Research Foundation (FAPESP).

Glossary

Activation function	Mathematical function that determines the output value of the layer or the node. Used to introduce nonlinearity to the neural networks.
Directed acyclic graph	Type of graph utilized in Bayesian networks aiming to detect the influence of variables has on the solution of a problem.
Hidden layers	Layer(s) between the input and output layers, where the nodes process the input and output signals weighting them accordingly their importance.
Hyperplane	It consists of some decision boundaries that help to classify the information into different classes or categories.
Nodes	The basic information processing unit in a neural network, where inputs coming from entered variables are processed and outputs are released from the node.
Split of dataset	It is the way in which the whole dataset is split into three subsets (i.e., training, validation, and test datasets).
Test	The last step of the training process. After the model is trained and validated, it is subjected to the test dataset (completely different sample from the training and validation data). This process aims to measure the final accuracy of trained AI, since this step is no more supervised as the training step.
Training	It is the first step to train an artificial neural network. It consists on the application of a sample (called training dataset), from which the machine will learn. The training dataset contains the features that characterize the

subject to be evaluated and the targets to be predicted.

Validation

The second step of the training process. Using the validation dataset, this step evaluates the model (i.e., the artificial neural network being trained) to ensure that AI training takes place correctly. With a different dataset from the training step, validation ensures that the model is not overfitted.

Weights

Weights are assigned to represent the relative importance of the input coming to the node.

References

1. Ledford H. IVF at 40: revisiting the revolution in assisted reproduction. *Nature*. 2018. <https://doi.org/10.1038/d41586-018-05792-9>.
2. Inhorn MC, Patrizio P. Infertility around the globe: new thinking on gender, reproductive technologies and global movements in the 21st century. *Hum Reprod Update*. 2014;21:411–26.
3. Vander Borgh M, Wyns C. Fertility and infertility: definition and epidemiology. *Clin Biochem*. 2018;62:2–10.
4. Begum MR. Assisted reproductive technology: techniques and limitations. *J Bangladesh Coll Physicians Surg*. 1970;26:135–41.
5. Younglai EV, Holloway AC, Foster WG. Environmental and occupational factors affecting fertility and IVF success. *Hum Reprod Update*. 2005;11:43–57.
6. Srouji SS, Mark A, Levine Z, Betensky RA, Hornstein MD, Ginsburg ES. Predicting in vitro fertilization live birth using stimulation day 6 estradiol, age, and follicle-stimulating hormone. *Fertil Steril*. 2005;84:795–7.
7. Ottosen LDM, Kesmodel U, Hindkjær J, Ingerslev HJ. Pregnancy prediction models and eSET criteria for IVF patients-do we need more information? *J Assist Reprod Genet*. 2007;24:29–36.
8. Al-Shawwa M, Abu-Naser SS. Predicting birth weight using artificial neural network. *Int J Acad Heal Med Res [Internet]*. 2019; 3:9–14. Available from: www.ijeais.org/ijahmr
9. Hota HS, Shukla SP, Kiran GK. Review of intelligent techniques applied for classification and preprocessing of medical image data. *Int J Comput Sci Issues*. 2013;10:267–72.
10. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019. <https://doi.org/10.1038/s41591-018-0307-0>.
11. Rich E, Knight K, Nair S. Artificial intelligence. 3rd ed. McGraw-Hill. 2008.
12. Peixoto MS, Barros LC, Bassanezi RC, Fernandes OA. An approach via fuzzy systems for dynamics and control of the soybean aphid. *Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology*. Atlantic Press. *Adv Intel Syst Res*. 2015;89:1295–301.
13. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. 1943;5:115–33.
14. Hsu F. IBM'S deep blue chess grandmaster chips. *IEEE Micro*. 1999;63:70–81.
15. Da Silva IN, Spatti DH, Flauzino RA, Liboni LHB, Dos Reis Alves SF. Artificial neural networks, vol. 39. Cham: Springer International Publishing; 2017.
16. Durairaj M, Thamilselvan P. Applications of artificial neural network for IVF data analysis and prediction. *J Eng Comput Appl Sci*. 2013;2:11–5.

17. Milewski R, Kuczyńska A, Stankiewicz B, Kuczyński W. How much information about embryo implantation potential is included in morphokinetic data? A prediction model based on artificial neural networks and principal component analysis. *Adv Med Sci*. 2017;62:202–6.
18. Girela JL, Gil D, Johnsson M, Gomez-Torres MJ, De Juan J. Semen parameters can be predicted from environmental factors and life-style using artificial intelligence methods. *Biol Reprod*. 2013;88:1–8.
19. El-Shafeiy E, El-Desouky A, El-Ghamrawy S. An optimized artificial neural network approach based on sperm whale optimization algorithm for predicting fertility quality. *Stud Inform Control*. 2018;27:349–58.
20. Sahoo AJ, Kumar Y. Seminal quality prediction using data mining methods. *Technol Health Care*. 2014;22:531–45.
21. Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. *Reprod Med Biol*. 2019;18:204–11.
22. Hassan MR, Al-Insaf S, Hossain MI, Kamruzzaman J. A machine learning approach for prediction of pregnancy outcome following IVF treatment. *Neural Comput & Applic*. 2018;32:2283–97. <https://doi.org/10.1007/s00521-018-3693-9>.
23. Manna C, Nanni L, Lumini A, Pappalardo S. Artificial intelligence techniques for embryo and oocyte classification. *Reprod BioMed Online*. 2013;26:42–9.
24. Candemir C. Estimating the semen quality from life style using fuzzy radial basis functions. *Int J Mach Learn Comput*. 2018;8:44–8.
25. Kumar V, Minz S. Feature selection: a literature review. *Smart Comput Rev*. 2014;4:211–29.
26. Ebrahimi A, Khamsehchi E. Sperm whale algorithm: an effective metaheuristic algorithm for production optimization problems. *J Nat Gas Sci Eng*. 2016;29:211–22. <https://doi.org/10.1016/j.jngse.2016.01.001>.
27. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2:359–66.
28. Riley P. Three pitfalls to avoid in machine learning. *Nature*. 2019;572:27–9.
29. 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.
30. Hernández-González J, Inza I, Crisol-Ortiz L, Guembe MA, Iñarra MJ, Lozano JA. Fitting the data from embryo implantation prediction: learning from label proportions. *Stat Methods Med Res*. 2018;27:1056–66.
31. Morales DA, Bengoetxea E, Larrañaga P. Selection of human embryos for transfer by Bayesian classifiers. *Comput Biol Med*. 2008;38:1177–86.
32. Corani G, Magli C, Giusti A, Gianaroli L, Gambardella LM. A Bayesian network model for predicting pregnancy after in vitro fertilization. *Comput Biol Med*. 2013. <https://doi.org/10.1016/j.compbiomed.2013.07.035>.
33. Morales DA, Bengoetxea E, Larrañaga P, García M, Franco Y, Fresnada M, et al. Bayesian classification for the selection of in vitro human embryos using morphological and clinical data. *Comput Methods Prog Biomed*. 2008;90:104–16.
34. 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 Mak*. 2015;35:714–25.
35. Jakkula V. Tutorial on support vector machine (SVM). *Sch EECS, Washington State Univ*. 2006;1–13.
36. Lima CADM. Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte. Repositorio da Produção Científica e Intelectual da Unicamp. 2004. <http://www.repositorio.unicamp.br/handle/REPOSIP/261258>.
37. Santos Filho E, Noble JA, Poli M, Griffiths T, Emerson G, Wells D. A method for semi-automatic grading of human blastocyst microscope images. *Hum Reprod*. 2012;27:2641–8.
38. Gardner D, Schoolcraft W. In-vitro culture of human blastocysts. In: Jansen R, Mortimer D, editors. *Towards reproductive certainty: infertility and genetics beyond*. Carnforth: Parthenon Press; 1999. p. 378–88.
39. Wang L-H, Fu Z-X, Ye S-Z, Ke D-G. Pattern recognition of IVF's early embryo images based on support vector machines and texture features. *Int J Eng Trends Technol*. 2018;66:7–11.
40. Xu L, Wei X, Yin Y, Wang W, Tian Y, Zhou M. Automatic classification of human embryo microscope images based on LBP feature. *Commun Comput Inf Sci*. 2014;437:145–52.
41. Pathak B, Barooah D. Texture analysis based on the gray-level co-occurrence matrix considering possible orientations. *Int J Adv Res Electr Electron Instrum Eng*. 2013;2:4206–12.
42. Rahim A, Hossain N, Wahid T, Azam S. Face recognition using local binary patterns (LBP). *Glob J Comput Sci Technol Graph Vis*. 2013;13:469–81.
43. Galloway C, Swiatek L. Public relations and artificial intelligence: it's not (just) about robots. *Public Relat Rev*. 2018;44:734–40.
44. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–43.
45. Goodfellow I, Bengio Y, Courville A. *Deep learning*. Cambridge: MIT Press; 2016.
46. Patil SN, Wali U, Swamy MK, Nagaraj SP, Patil N. Deep learning techniques for automatic classification and analysis of human in vitro fertilized (IVF) embryos. *J Emerg Technol Innov Res*. 2018;5:100–6.
47. Khan A, Gould S, Salzmann M. Segmentation of developing human embryo in time-lapse microscopy. *International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2016; IEEE, TBC;930–934*.
48. Chen T-J, Zheng W-L, Liu C-H, Huang I, Lai H-H, Liu M. Using deep learning with large dataset of microscope images to develop an automated embryo grading system. *Fertil Reprod*. 2019;1:51–6.
49. Matusevičius A, Dirvanauskas D, Maskeliūnas R, Raudonis V. Embryo cell detection using regions with convolutional neural networks. *CEUR Workshop Proc*. 1856;2017:89–93.
50. Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of deep learning for predicting live birth from a blastocyst image in patients classified by age. *Reprod Med Biol*. 2019;18:190–203.
51. Dirvanauskas D, Maskeliūnas R, Raudonis V, Damasevicius R. Embryo development stage prediction algorithm for automated time lapse incubators. *Comput Methods Prog Biomed*. 2019;177:161–74. <https://doi.org/10.1016/j.cmpb.2019.05.027>.
52. Cao Q, Liao SS, Meng X, Ye H, Yan Z, Wang P. Identification of viable embryos using deep learning for medical image. *ACM 5th Int Conf Proceeding Ser*. 2018;69–72.
53. Khosravi P, Kazemi E, Zhan Q, Malmsten JE, Toschi M, Zisimopoulos P, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digital Med*. 2019. <https://doi.org/10.1038/s41746-019-0096-y>.
54. Rad RM, Saeedi P, Au J, Havelock J. Blastomere cell counting and centroid localization in microscopic images of human embryo. *2018 IEEE 20th Int Work Multimed Signal Process MMSP 2018*. IEEE; 2018;1–6.
55. Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod*. 2019;34:1011–8.
56. Bormann CL, Thirumalaraju P, Kanakasabapathy MK, Kandula H, Souter I, Dimitriadis I, et al. Consistency and objectivity of automated embryo assessments using deep neural networks. *Fertil Steril*. 2020;113:781–7.

57. VerMilyea M, Hall JMM, Diakiw SM, Johnston A, Nguyen T, Perugini D, et al. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod.* 2020;35:770–84.
58. Kragh MF, Rimestad J, Berntsen J, Karstoft H. Automatic grading of human blastocysts from time-lapse imaging. *Comput Biol Med.* 2019;115:103494. <https://doi.org/10.1016/j.combiomed.2019.103494>.
59. Ratna MB, Bhattacharya S, Abdulrahim B, McLernon DJ. A systematic review of the quality of clinical prediction models in in vitro fertilisation. *Hum Reprod.* 2020;35(1):100–16.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.