

New Methods for Comparing Embryo Selection Methods by Applying Artificial Intelligence

Yasunari Miyagi

Medical Data Labo, Okayama, Japan
Department of Gynecologic Oncology,
Saitama Medical University International Medical Center
Saitama, Japan

Toshihiro Habara

Okayama Couple's Clinic, Okayama, Japan

Rei Hirata

Okayama Couple's Clinic, Okayama, Japan

Nobuyoshi Hayashi

Okayama Couple's Clinic, Okayama, Japan

ABSTRACT

Purpose: To develop a process for comparing embryo selection methods by applying artificial intelligence (AI) for live births retrospectively with historical individual embryo sequence data from live births. **Methods:** We sought to predict the results for live birth outcomes with the sequences of embryos from patient data. To do so, we developed two functions. For the first (ranking evaluation) method, we summed the weighted order number of the embryo sequence to determine the order number of embryos that became a live birth. For the second (probability evaluation) method, we calculated the AI-derived confidence scores of the embryo for predicting a live birth via the result of the embryo sequence. **Results:** We used both methods to evaluate the embryo sequence for empirical samples. With the assumed known result of becoming a live birth in a sequence of blastocysts, any methods for creating sequences could now be quantitatively evaluated, regardless of whether AI is used. **Conclusions:** Through use of AI, we could, to a certain extent, expect to evaluate the superiority or inferiority of past, current, or future methods. The accuracy of AI-based diagnosis for predicting a live birth should improve, meaning the methods presented herein will become increasingly useful.

Keywords: Artificial Intelligence, Blastocyst, Deep Learning, Live Birth, Computational Neural Network.

INTRODUCTION

The goal of reproductive medicine is to help individuals achieve live births. In the field of assisted reproductive technology, several scoring systems for morphological evaluation of embryos that are to be transferred to the uterus have been investigated [1-3]; these include the Gardner blastocyst grading system [4]. Selected embryos from the patient are transferred in

accordance with morphological evaluation methods, but assurance of a live birth has not been established. Recently, attempts to use artificial intelligence (AI) to predict live birth achievement in reproductive medicine [5-12] have also been reported, similar to in other clinical studies [13-20]. There are AI systems that use not only images as morphological data but also conventional embryo evaluation parameters (e.g., patient age, body mass index, anti-Müllerian hormone level) as the non-morphological data [5]. With AI classifiers, prediction of live birth achievement can be displayed as whether the embryo will become a live birth by a classifier, or as a confidence score that is the AI-generated numerical probability of becoming a live birth. Notably, the probability-based evaluation can be used to sequentially order the embryos. Use of highly accurate AI is expected to become an important assistive tool in diagnosis; therefore, to compare embryo evaluation methods including many types of AI and the morphological evaluation methods, it is desirable to compare them retrospectively. This can be based on data in which it is known whether a live birth was achieved, because clinical randomized trials to comparing more than two evaluation methods are considered impossible for embryo transfer. Moreover, when an embryo is transferred and a live birth is achieved, an embryo being stored will not be used until the next time a birth is desired. It is unclear whether non-transferred embryos will become live births. It is therefore not possible to accurately compare methods because there are embryos that may or may not result in live births.

The current conventional superiority indices for various embryo evaluation methods, including AI, are accuracy, sensitivity, specificity, and the value of the area under the curve [21]. These generally are quite normal indices for evaluating a certain test method on a target population. When considering each patient as the main subject in the clinical practice of reproductive medicine, however, it is important to ascertain the superiority or inferiority of the patient's multiple embryos and how to select the best embryo. Evaluation results on other people's embryos are irrelevant to the patient.

Selecting the best method for ranking embryos for each patient is therefore required by comparing methods for ranking embryos using embryos with uncertain past information, retrospectively. We developed two methods to retrospectively infer comparison of different embryo evaluation methods to some extent using data on patients who obtained a live birth by single embryo transfer from multiple embryos. The two methods are demonstrated with using empirical samples. Our methods can be used for selecting the best embryo by choosing the best among many past, current, or future evaluation methods.

MATERIALS AND METHODS

In an embryo evaluation method, when a is an embryo, the order of transferred embryos in the order of superiority in a group of N embryos is as follows:

$$a_i | 0 < i \leq N, i \in \mathbb{N},$$

In this group, suppose embryos are implanted starting from $i=1$, and a live birth is obtained when $i=m$ ($0 < m \leq N$, $m \in \mathbb{N}$). It is a fact that no embryo was a live birth for $a_k | k < m$ ($0 < k \leq N$, $k \in \mathbb{N}$) and it is unknown whether the embryo became a live birth for $a_k | k > m$.

Ranking evaluation (non-parametric) method

We define function f as follows:

$$f(m, w) = \sum_{i=1}^m i + \sum_{i=m+1}^N (i + w)$$

where w ; weight, $w \in \mathbb{R}$

If we calculate $f(m, w)$ in the permutation of a_i , we can determine that the smaller is $f(m, w)$, the better is the order of administration (see Suppl. 1, in which the blastocyst that became a live birth is shown in red and with an asterisk).

Suppose four simulated embryos were enrolled. The embryo grading order by a method (method A) was thought to be a_1, a_2, a_3 , and a_4 , where a_1 was the best by definition. After the embryos were transferred with the sequence following method A, it is supposed that a_1 and a_2 were failures and the third embryo became the first live birth. Then suppose a classifier (classifier B) by AI shows the best sequence as a_4, a_3, a_1 , and a_2 , which means the second embryo would become a live birth, and another classifier (classifier C) by AI shows the best sequence as a_2, a_1, a_4 , and a_3 , which means the fourth embryo would become a live birth. We compared classifier B and C using the values of $f(m, w)$.

Probability evaluation (parametric) method

Let p_i be the probability of live birth achievement calculated by AI at each a_i and $p = \{p_1, p_2, \dots, p_N\}$. Let $s[m]$ be the ranking sequence number of a_m .

We define function g as follows:

$$g(m, p) = \sum_{n=1}^N \{R(p_n) \times \prod_{k=1}^{n-1} (1 - p_k)\} / s[m]$$

$$R(p_x) = \begin{cases} p_x & \text{if } x \geq m \\ 0 & \text{if } x < m \end{cases}$$

If we calculate $g(m, p)$ in the permutation of p_i , we can judge that the larger is $g(m, p)$, the better is the order of administration (see Suppl. 2, in which the blastocyst that became a live birth is shown in red and with asterisk).

Suppose four simulated embryos were enrolled and the embryo grading order by method A was thought be a_1, a_2, a_3 , and a_4 . After the embryos were transferred following the method, it is supposed that a_1 and a_2 were failures and the third embryo became the first live birth. For an empirical case, suppose the confidence score that is a probability for achieving a live birth by classifier D is 0.60, 0.55, 0.50, and 0.45 for a_1, a_2, a_3 , and a_4 , respectively. For another empirical case, suppose the confidence score that is a probability for achieving a live birth by classifier E is 0.40, 0.70, 0.20, and 0.60 for a_1, a_2, a_3 , and a_4 , respectively. For an additional empirical case, suppose the confidence score that is a probability for achieving a live birth by classifier F is 0.40, 0.70, 0.80, and 0.60 for a_1, a_2, a_3 , and a_4 , respectively. We compared classifiers D, E, and F using the values of the $g(m, p)$.

RESULTS

Ranking evaluation method (Table 1)

Classifier A gives $f(3,w)=1+2+3=6$ for a sequence of a_i . Classifier B gives $f(3,w)=1+w+2=3+w$ for a sequence of a_i . Classifier C gives $f(3,w)=1+2+3+w+4=10+w$ for a sequence of a_i . Then, if w is 0.5 that is half the ranking interval, $f(3,0.5)$ of classifiers A, B, and C is 6, 3.5, and 10.5, respectively. Classifier B therefore appears to be the best.

Table 1. We suppose the blastocyst quality decided by a method is $a[1]$, $a[2]$, $a[3]$, and $a[4]$ in descending order by which the embryo would be transferred. The formula of $f(m,w)$ of the ranking evaluation method for the sequences of the four blastocysts is shown. An asterisk indicates a blastocyst that became a live birth. Refer to Suppl. 1 for further specifics

sequence				$f(3,w)$	$f(3,0.5)$
$a[1]$	$a[2]$	$a[3]^*$	$a[4]$	6	6
$a[1]$	$a[2]$	$a[4]$	$a[3]^*$	$10+w$	10.5
$a[1]$	$a[3]^*$	$a[2]$	$a[4]$	3	3
$a[1]$	$a[3]^*$	$a[4]$	$a[2]$	3	3
$a[1]$	$a[4]$	$a[2]$	$a[3]^*$	$10+w$	10.5
$a[1]$	$a[4]$	$a[3]^*$	$a[2]$	$6+w$	6.5
$a[2]$	$a[1]$	$a[3]^*$	$a[4]$	6	6
$a[2]$	$a[1]$	$a[4]$	$a[3]^*$	$10+w$	10.5
$a[2]$	$a[3]^*$	$a[1]$	$a[4]$	3	3
$a[2]$	$a[3]^*$	$a[4]$	$a[1]$	3	3
$a[2]$	$a[4]$	$a[1]$	$a[3]^*$	$10+w$	10.5
$a[2]$	$a[4]$	$a[3]^*$	$a[1]$	$6+w$	6.5
$a[3]^*$	$a[1]$	$a[2]$	$a[4]$	1	1
$a[3]^*$	$a[1]$	$a[4]$	$a[2]$	1	1
$a[3]^*$	$a[2]$	$a[1]$	$a[4]$	1	1
$a[3]^*$	$a[2]$	$a[4]$	$a[1]$	1	1
$a[3]^*$	$a[4]$	$a[1]$	$a[2]$	1	1
$a[3]^*$	$a[4]$	$a[2]$	$a[1]$	1	1
$a[4]$	$a[1]$	$a[2]$	$a[3]^*$	$10+w$	10.5
$a[4]$	$a[1]$	$a[3]^*$	$a[2]$	$6+w$	6.5
$a[4]$	$a[2]$	$a[1]$	$a[3]^*$	$10+w$	10.5
$a[4]$	$a[2]$	$a[3]^*$	$a[1]$	$6+w$	6.5
$a[4]$	$a[3]^*$	$a[1]$	$a[2]$	$3+w$	3.5
$a[4]$	$a[3]^*$	$a[2]$	$a[1]$	$3+w$	3.5

Probability evaluation method (Table 2)

The modified confidence score of classifier D will then be 0, 0, 0.50, and 0.45, respectively. The value of $g(3, \{0, 0, 0.50, 0.45\})$ by classifier D is then $\{0 + (1-0)*0 + (1-0)*(1-0)*0.50\}/3 = 0.1667$. The modified confidence score of classifier E will be 0, 0, 0.20, and 0.60, respectively. The value of $g(4, \{0, 0, 0.20, 0.60\})$ by classifier E is then $\{0 + (1-0)*0.6 + (1-0)*(1-0.6)*0 + (1-0)*(1-0.6)*(1-0)*0.20\}/4 = 0.17$ for a sequence of a_2, a_4, a_1 , and a_3 in descending order. The modified confidence score of classifier F will be 0, 0, 0.80, and 0.60, respectively. The value of $g(1, \{0, 0, 0.80, 0.60\})$ by classifier F is $\{0.8 + (1-0.8)*0 + (1-0.8)*(1-0)*0.6 + (1-0.8)*(1-0)*(1-0.6)*0\}/1 = 0.92$ for a sequence of a_3, a_2, a_4 , and a_1 in descending order. Classifier F therefore appears to be the best.

Table 2. We suppose the blastocyst quality decided by a method is $a[1], a[2], a[3]$, and $a[4]$ in descending order by which the embryo would be transferred. The probabilities of achieving a live birth, which are the confidence scores calculated by artificial intelligence classifiers, of $a[1], a[2], a[3]$, and $a[4]$ are p_1, p_2, p_3 , and p_4 , respectively, $q_i = 1 - p_i$. Formula of $g(m, p)$, p ; the sequence of the confidence scores, of the probability evaluation method for the four blastocysts' sequences are shown. An asterisk indicates a blastocyst that became a live birth.

Refer to Suppl. 2 for further specifics

sequence				m	$g(m, p)$
$a[1]$	$a[2]$	$a[3]^*$	$a[4]$	3	$1/3 (p_3 q_1 q_2 + p_4 q_1 q_2 q_3)$
$a[1]$	$a[2]$	$a[4]$	$a[3]^*$	4	$1/4 (p_4 q_1 q_2 + p_3 q_1 q_2 q_4)$
$a[1]$	$a[3]^*$	$a[2]$	$a[4]$	2	$1/2 (p_3 q_1 + p_4 q_1 q_2 q_3)$
$a[1]$	$a[3]^*$	$a[4]$	$a[2]$	2	$1/2 (p_3 q_1 + p_4 q_1 q_3)$
$a[1]$	$a[4]$	$a[2]$	$a[3]^*$	4	$1/4 (p_4 q_1 + p_3 q_1 q_2 q_4)$
$a[1]$	$a[4]$	$a[3]^*$	$a[2]$	3	$1/3 (p_4 q_1 + p_3 q_1 q_4)$
$a[2]$	$a[1]$	$a[3]^*$	$a[4]$	3	$1/3 (p_3 q_1 q_2 + p_4 q_1 q_2 q_3)$
$a[2]$	$a[1]$	$a[4]$	$a[3]^*$	4	$1/4 (p_4 q_1 q_2 + p_3 q_1 q_2 q_4)$
$a[2]$	$a[3]^*$	$a[1]$	$a[4]$	2	$1/2 (p_3 q_2 + p_4 q_1 q_2 q_3)$
$a[2]$	$a[3]^*$	$a[4]$	$a[1]$	2	$1/2 (p_3 q_2 + p_4 q_2 q_3)$
$a[2]$	$a[4]$	$a[1]$	$a[3]^*$	4	$1/4 (p_4 q_2 + p_3 q_1 q_2 q_4)$
$a[2]$	$a[4]$	$a[3]^*$	$a[1]$	3	$1/3 (p_4 q_2 + p_3 q_2 q_4)$
$a[3]^*$	$a[1]$	$a[2]$	$a[4]$	1	$p_3 + p_4 q_1 q_2 q_3$
$a[3]^*$	$a[1]$	$a[4]$	$a[2]$	1	$p_3 + p_4 q_1 q_3$
$a[3]^*$	$a[2]$	$a[1]$	$a[4]$	1	$p_3 + p_4 q_1 q_2 q_3$
$a[3]^*$	$a[2]$	$a[4]$	$a[1]$	1	$p_3 + p_4 q_2 q_3$
$a[3]^*$	$a[4]$	$a[1]$	$a[2]$	1	$p_3 + p_4 q_3$
$a[3]^*$	$a[4]$	$a[2]$	$a[1]$	1	$p_3 + p_4 q_3$
$a[4]$	$a[1]$	$a[2]$	$a[3]^*$	4	$1/4 (p_4 + p_3 q_1 q_2 q_4)$

$a[4]$	$a[1]$	$a[3]^*$	$a[2]$	3	$1/3 (p_4+p_3 q_1 q_4)$
$a[4]$	$a[2]$	$a[1]$	$a[3]^*$	4	$1/4 (p_4+p_3 q_1 q_2 q_4)$
$a[4]$	$a[2]$	$a[3]^*$	$a[1]$	3	$1/3 (p_4+p_3 q_2 q_4)$
$a[4]$	$a[3]^*$	$a[1]$	$a[2]$	2	$1/2 (p_4+p_3 q_4)$
$a[4]$	$a[3]^*$	$a[2]$	$a[1]$	2	$1/2 (p_4+p_3 q_4)$

DISCUSSION

We developed two methods for comparing evaluation methods of embryo evaluation predictions based on embryo transfer results involving supposed non-transferred embryos. We used historical individual embryo sequence data from the live birth to compare several classification methods that are the past, current, and/or future morphological evaluation methods and past, current, and/or future methods using AI. An empirical sample for the ranking evaluation method shows the performances of the AI-classifiers could be compared using the sequences of a patient's embryo order. The classifier's performance could be expressed using the value of $f(m,w)$ with known results. An empirical sample for the probability evaluation method shows the AI classifiers' performance could be compared using confidence scores with the sequences. For classifier F, which has 0.40, 0.70, 0.80, and 0.60 as confidence scores for a_1 , a_2 , a_3 , and a_4 , respectively, the transfer sequence would be a_3 , a_2 , a_4 , and a_1 , if the classifier was used, resulting in 0.92 of $g(m, \{p_3, p_2, p_4, p_1\})$. If the transfer sequence is a_1 , a_2 , a_3 , and a_4 , the value of $g(m, \{p_1, p_2, p_3, p_4\})$ would be 0.307 by calculation. We were able to apply our methods to quantitatively demonstrate the importance of the transfer sequence.

We applied our methods for evaluating a live birth, but they can also theoretically be used for evaluating implantation. The target evaluation method can either be a ranking evaluation method in which no numerical value is given for each embryo or a probability evaluation method in which a numerical value is given. The former evaluates the order of embryo selection itself in non-parametric procedures, while the latter evaluates the expected value of the probability owing to the order in parametric procedures.

It is theoretically possible to create a wide variety of AI classifiers potentially capable of predicting live births. The AI classifiers could calculate each embryo's probability of a live birth. We have previously reported methods of calculating the probability of a live birth for each blastocyst, either from a single image or by combining images and conventional embryo evaluation parameters [5-7]. Methods have also been reported that use machine learning to evaluate embryos [11,22] from multiple images, such as time lapse [11,23-25].

Irrespective of what information is input, the output will be a classification of each embryo as a live birth or not a live birth, and/or a numerical value based on the probability of a live birth. In clinical reproductive medicine, however, where randomized trials are considered impossible, there is no authoritative way to determine the optimal method of embryo evaluation. With the two methods presented herein, we can expect to determine, to some extent, the superiority or inferiority of past, current, and future evaluation methods. Even if many classification methods, including AI, emerge in the future, we expect our methods will permit a certain extent of evaluation.

There are some limitations to our methods. The first limitation is that the conclusions our methods reach are inevitably speculative, and definitive results cannot be assured. This is unavoidable because the methods involve assumed expectation of unknown embryos. Additionally, for the ranking evaluation method, the maximum likelihood optimal value of the value of weight (w) as a penalty in the selection order for unknown embryos is not known. If $w=0$, there is no penalty and it is considered too light, while if $w=1$, the ranking is equal to the next case and it is considered too heavy; thus, $w=0.5$ is considered theoretically reasonable. The probability evaluation method also cannot be applied where the embryo selection method is not based on numerical values, such as with the Gardner blastocyst grading system. In other words, it is not possible to directly compare recent AI methods with the Gardner system. There is, however, still a possibility of comparing the Gardner system by using our multivariate logistic function [5-7], created with reference to the Gardner system. Additionally, regarding comparison of the ranking and probability evaluation methods, for the former, if the embryo order is the same in the two embryo selection methods, the methods are judged to have the same performance. In the latter case, however, even if the embryo order in the two selection methods is the same, if the expected probability of live birth for each embryo differs, the order will show a different expected probability. If the results of the selection method evaluation are not consistent between the ranking and probability evaluation methods, the judgment of superiority or inferiority must be subjective at the present because the calculation methods are independent, and they differ for each method. A further limitation is that, ideally, our methods should be validated. A perfect validation study, however, was impossible because the outcomes of non-transferred embryos are unknown. Finally, from a statistical standpoint, because each of our methods was applied to a group of embryos in a patient, the methods need to be further applied to many patients who achieved a live birth in order to compare and select the best method for embryo evaluation prediction.

Regarding prediction of live birth achievement, we should, in the future, move toward having AI make decisions using both images and non-images. Images may be a single image, multiple images at the same time for each microscopically observed depth, and/or multiple images in a time series, such as a time lapse. Non-images may be, for example, morphological and functional information on the uterus, bacterial flora, immunological information, or genetic information on the embryo, in addition to the conventional embryo evaluation parameters. Thus, as the accuracy of AI-based diagnosis is expected to improve, our methods presented herein will be useful in selecting the AI.

The methods in this study appear applicable for efficiently selecting the best embryo by choosing the best among past, current, and future methods, which likely will use AI.

References

- Dennis SJ, Thomas MA, Williams DB, Robins JC. Embryo morphology score on day 3 is predictive of implantation and live birth rates. *J. Assist. Reprod Genet.* 2006; 23: 171–75.
- Goto S, Kadowaki T, Tanaka S, Hashimoto H, Koikeguchi S, Shiotani M. Prediction of pregnancy rate by blastocyst morphological score and age, based on 1,488 single frozen-thawed blastocyst transfer cycles. *Fertil Steril.* 2011; 95: 948–52.
- Van den Abbeel E, Balaban B, Ziebe S, Lundin K, Cuesta MJ, Klein BM, et al. Association between blastocyst morphology and outcome of single-blastocyst transfer. *Reprod Biomed Online.* 2013; 27: 353–61.
- Gardner DK, Lane M, Stevens J, Schlenker T, Schoolcraft WB. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertil Steril.* 2000 Jun; 73 :1155–8.
- Miyagi Y, Habara T, Hirata R, Hayashi N. Predicting a live birth by artificial intelligence incorporating both the blastocyst image and conventional embryo evaluation parameters. *Artif Intell Med Imaging.* 2020; 1: 94–107.
- 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.
- Miyagi Y, Habara T, Hirata R, Hayashi N. Feasibility of predicting live birth by combining conventional embryo evaluation with artificial intelligence applied to a blastocyst image in patients classified by age. *Reprod Med Biol.* 2019; 18: 344–56.
- 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.
- Miyagi Y, Habara T, Hirata R, Hayashi N. Deep Learning to predicting live births and aneuploid miscarriages from images of blastocysts combined with maternal age. *Int J Bioinform Intell Comput.* 2022 in printing.
- 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 Digit Med.* 2019; 2: 21.
- 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.
- Chavez-Badiola A, Flores-Saiffe Farias A, Mendizabal-Ruiz G, Garcia-Sanchez R, Drakeley AJ, Garcia-Sandoval JP. Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. *Sci Rep.* 2020; 10: 4394.
- Miyagi Y, Hata T, Bouno S, Koyanagi A, Miyake T. Recognition of facial expression of fetuses by artificial intelligence (AI). *J Perinat Med.* 2021; 49: 596–603.
- Miyagi Y, Hata T, Bouno S, Koyanagi A, Miyake T. Recognition of fetal facial expressions using artificial intelligence deep learning. *Donald School J Ultrasound Obstet Gynecol.* 2021; 15: 223–8.
- Miyagi Y, Miyake T. Potential of artificial intelligence for estimating Japanese fetal weights. *Acta Medica Okayama.* 2020; 74: 483–93.
- Miyagi Y, Takehara K, Nagayasu Y, Miyake T. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images combined with HPV types. *Oncol Lett.* 2020; 19: 1602–10.
- Miyagi Y, Takehara K, Miyake T. Application of deep learning to the classification of uterine cervical squamous epithelial lesion from colposcopy images. *Mol Clin Oncol.* 2019; 11: 583–89.
- Miyagi Y, Tada K, Yasuhi I, Maekawa Y, Okura N, Kawakami K, et al. New method for determining fibrinogen and FDP threshold criteria by artificial intelligence in cases of massive hemorrhage during delivery. *J Obstet Gynaecol Res.* 2020; 46: 256–65.
- Miyagi Y, Fujiwara K, Oda T, Miyake T, Coleman RL. Studies on development of new method for the prediction of clinical trial results using compressive sensing of artificial intelligence. In: Ferreira MAM, editor. *Theory and practice of mathematics and computer science.* Hooghly: Book Publisher International; 2020, p. 101–8.
- Miyagi Y, Fujiwara K, Oda T, Miyake T, Coleman RL. Development of new method for the prediction of clinical trial results using compressive sensing of artificial intelligence. *J Biostat Biometric App.* 2018; 3: 202.

Kragh MF, Karstoft H. Embryo selection with artificial intelligence: how to evaluate and compare methods? J Assist Reprod Genet. 2021; 38: 1675–89.

Chavez-Badiola A, Farias AFS, Mendizabal-Ruiz G, Garcia-Sanchez R, Drakeley AJ, Garcia-Sandoval JP. Predicting pregnancy test results after embryo transfer by image feature extraction and analysis using machine learning. 2020; Sci Rep. 10: 1–6.

Abbasi M, Saeedi P, Au J, Havelock, J. A deep learning approach for prediction of IVF implantation outcome from day 3 and day 5 time-lapse human embryo image sequences. 2021 IEEE International Conference on Image Processing (ICIP); 2021 Sep 19-22, p. 289–93

Kragh MF, Rimestad J, Berntsen J, Karstoft H. Automatic grading of human blastocysts from time-lapse imaging. Comput Biol Med. 2019; 115: 103494.

Sawada Y, Sato T, Nagaya M, Saito C, Yoshihara H, Banno C, et al. Evaluation of artificial intelligence using time-lapse images of IVF embryos to predict live birth. Reprod Biomed Online. 2021; 43: 843–52.