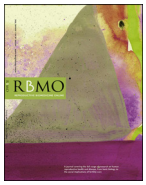
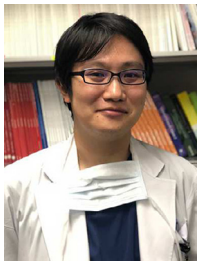
RBMO

ELSEVIER



ARTICLE

Improved pregnancy prediction performance in an updated deep-learning embryo selection model: a retrospective independent validation study



BIOGRAPHY

Dr Satoshi Ueno is Senior Embryologist at the Kato Ladies Clinic in Tokyo, Japan. Dr Ueno earned his PhD from Okayama University. His main research interests are intracytoplasmic sperm injection, gametes cryopreservation, embryo morphokinetics, embryo selection for transfer and vitrified–warmed blastocyst transfer.

Satoshi Ueno[a], Jørgen Berntsen[b], Tadashi Okimura[a], Keiichi Kato[a,*]

KEY MESSAGE

The area under curve for the prediction of ongoing pregnancy was significantly higher in a deep-learning model with 15% more training data than its previous version and than Gardner grading. Continuous collection of training data is an important step for improving the performance of updated deep-learning models for pregnancy prediction.

ABSTRACT

Research question: What is the effect of increasing training data on the performance of ongoing pregnancy prediction after single vitrified–warmed blastocyst transfer (SVBT) in a deep-learning model?

Design: A total of 3960 SVBT cycles were retrospectively analysed. Embryos were stratified according to the Society for Assisted Reproductive Technology age groups. Embryos were scored by deep-learning models iDAScore v1.0 (IDA-V1) and iDAScore v2.0 (IDA-V2) (15% more training data than v1.0) and by Gardner grading. The discriminative performance of the pregnancy prediction for each embryo scoring model was compared using the area under the curve (AUC) of the receiver operating characteristic curve for each maternal age group.

Results: The AUC of iDA-V2, iDA-V1 and Gardener grading in all cohort were 0.736, 0.720 and 0.702, respectively. iDA-V2 was significantly higher than iDA-V1 and Gardener grading ($P < 0.0001$). Group > 35 years ($n = 757$): the AUC of iDA-V2 was significantly higher than Gardener grading (0.718 versus 0.694, $P = 0.015$); group aged 35–37 years ($n = 821$), the AUC of iDA-V2 was significantly higher than iDA-V1 (0.712 versus 0.696, $P = 0.035$); group aged 41–42 years ($n = 715$, the AUC of iDA-V2 was significantly higher than Gardener grading (0.745 versus 0.696, $P = 0.007$); group > 42 years ($n = 660$) and group aged 38–40 years ($n = 1007$), no significant differences were found between the groups.

Conclusion: The performance of deep learning models for pregnancy prediction will be improved by increasing the size of the training data.

[a] Kato Ladies Clinic, 7-20-3, Nishi-shinjuku, Shinjuku, Tokyo 160-0023, Japan.
[b] Vitrolife A/S, Jens Juuls Vej 20 8260 Viby J, Denmark.

*Corresponding author. E-mail address: k-kato@towako.net (K. Kato). https://doi.org/10.1016/j.rbmo.2023.103308 1472-6483/
Declaration: JB is an employee and shareholder of Vitrolife A/S.

KEYWORDS

pregnancy prediction
single cryopreserved blastocyst transfer
artificial intelligence
deep learning model
blastocyst assessment

INTRODUCTION

Artificial intelligence technology is increasingly used for diagnostic purposes in human assisted reproductive technology (ART) (*Zaninovic and Rosenwaks, 2020*). Extensive studies have been conducted on the performance of artificial intelligence for pregnancy prediction after blastocyst transfer (*Zhan et al., 2020*; *Ueno et al., 2021*; *Glatstein et al., 2023*; in press). A benefit of methods based on artificial intelligence is that they are objective, consistent, standardized, and efficient tools for embryo evaluation (*Zaninovic and Rosenwaks, 2020*). Additionally, the accuracy of artificial intelligence models, especially the deep-learning model, will improve with increasing the training dataset (*Xue, 2019*).

The iDAScore® version 1.0 model (Vitrolife, Gothenburg, Sweden) is a commercial deep-learning model for scoring blastocysts. Several studies have shown that iDAScore version 1.0 (iDA-V1) (Vitrolife, Gothenburg, Sweden) is significantly correlated with implantation, ongoing pregnancy, miscarriage, live birth and euploid rates (*Zhan et al., 2020*; *Ueno et al., 2021*; *Glatstein et al., 2023*; *Kato et al., 2023*; in press). Recently, iDAScore version 2.0 (iDA-V2; Vitrolife, Sweden) was launched. iDA-V2 and iDA-V1 were based on an identical deep-learning architecture, but the training data for iDA-V2 has been increased by 15% (*Lassen et al., 2023*).

The performance of embryo selection models increases with a larger training data set. For example, the performance of an embryo selection model based on artificial intelligence increased from 0.5 to 0.67 when the number of transferred embryos increased from 416 to 9871 (*Kragh* et al., *2022*). This study, however, was conducted on internal validation data and on models that are not publicly available. To our knowledge, no external validation studies have been published on the effect of increasing training data on the performance of deep learning models for pregnancy prediction.

In the present study, the performance of pregnancy prediction, defined as a positive fetal heartbeat, was compared between iDA-V2 and iDA-V1 (or Gardner grading, a general blastocyst grading system) using the area under the curve (AUC) from receiver operating characteristic (ROC) analysis for different age groups after single vitrified–warmed blastocyst transfer (SVBT). Additionally, the influence of updating the iDAScore on the correlation between iDA-V2 and blastocyst morphological grade, the day of blastocyst vitrification and the grade of blastocyst expansion was analysed.

MATERIALS AND METHODS

Patients and study design

The records of 3960 autologous SVBT cycles from 3960 patients who underwent their first SVBT cycle in our clinic between April 2021 and May 2022 (one patient, one cycle) were reviewed. Cycles involving preimplantation genetic testing and cycles without a complete time-lapse sequence (owing to instrument maintenance) were excluded from the study. Furthermore, cycles in which embryos were cultured for less than 112 h after insemination were excluded because iDAScore can only be calculated for embryos cultured for 112 h or more. Single vitrified–warmed blastocyst transfer was carried out on days 4.5–5 after ovulation during a natural or hormone replacement cycle (*Takeshima et al., 2022*). iDA-V2 and iDA-V1 have been developed on a large dataset from multiple IVF clinics, including our clinic. Data used in the present study were derived from treatments carried out after the training phase and should be considered as independent data called temporal external validation data by TRIPOD guideline (*Moons et al., 2015*). During the study period, our centre strictly adhered to an exclusive single-embryo transfer policy and, therefore, only single-embryo transfers were carried out.

The cycles were stratified into five maternal age groups according to the Society for Assisted Reproductive Technology age groups (<35, 35–37, 38–40, 41–42 and >42 years).

The main performance measurement in this study was the AUC for discrimination of predicting an ongoing pregnancy (confirmed fetal heartbeat at 9 weeks of pregnancy) within each maternal age group for iDA-V2, iDA-V1 and Gardner grading. Second, the correlation between iDA-V2 or iDA-V1 and static blastocyst characteristics was analysed.

Ethical approval

The present retrospective study was reviewed and approved by the independent Institutional Review Board of Kato Ladies Clinic, Tokyo (approval number: 18–20, 20 August 2020). Written informed consent was obtained from all couples, who were informed that their de-identified data could be used for retrospective analyses.

Minimal ovarian stimulation, oocyte retrieval, fertilization procedures and embryo culture

All patients underwent minimal ovarian stimulation. A detailed protocol for minimal stimulation with clomiphene citrate has been previously reported (*Teramoto and Kato, 2007*; *Kato et al., 2012*). Briefly, clomiphene citrate (50 mg/day) (Fuji Pharma Co., Ltd., Tokyo, Japan) was orally administered with an extended regimen, starting on day 3 of the retrieval cycle to the day before triggering of final oocyte maturation. Monitoring, which involved an ultrasound scan and hormonal profiles, was usually initiated on day 8 and was carried out continuously every other day until the day of ovulation trigger. A nasal spray containing the gonadotrophin releasing hormone agonist buserelin (Suprecur) (Mochida Pharmaceutical Co., Ltd., Tokyo, Japan or Buserecur, Fuji Pharma Co., Ltd.) was used for ovulation trigger. A minimal dose of human menopausal gonadotrophin (HMG) (Ferring Pharma Co., Ltd., Saint-Prex, Switzerland) or recombinant FSH (Merck & Co., Rahaway, NJ, USA) was administered on days 8 and 10 to induce final follicular growth and maturation when more than three follicles were of uniform size (smaller than 10 mm) in the ovary on day 8, and the serum FSH level was less than 15 mIU/ml. When the serum FSH level was less than 10 mIU/ml, 150 IU of either HMG or recombinant FSH was administered. When the serum FSH level on day 8 was 10–15 mIU/ ml, 75 IU of HMG or recombinant FSH was administered. Oocyte retrieval was usually carried out 30–36 h after triggering using a 21-G needle (Kitazato Corporation, Shizuoka, Japan), generally without anaesthesia or follicular flushing. Cumulus–oocyte complexes were collected, washed and then transferred to human tubal fluid medium (Kitazato Corporation, Shizuoka, Japan) with paraffin oil at 5% CO_2 in air at 37°C. Conventional insemination was carried out approximately 3 h after retrieval, and intracytoplasmic sperm injection (ICSI) was carried out 45 h after retrieval. Approximately 5 h after conventional insemination, oocytes were transferred to a pre-equilibrated EmbryoSlide (Vitrolife,

Gothenburg, Sweden) and incubated in a time-lapse incubator (EmbryoScope$^+$, EmbryoScope Flex) (Vitrolife, Gothenburg, Sweden). Oocytes inseminated by ICSI were transferred immediately after ICSI to a pre-equilibrated EmbryoSlide and incubated in a time-lapse incubator (EmbryoScope$^+$; EmbryoScope Flex) (Vitrolife, Gothenburg, Sweden). Culture dishes were prepared according to the manufacturer's instructions. A one-step medium (NAKA Medical, Tokyo, Japan) was used for embryo culture. The culture dishes were covered with mineral oil (Ovoil) (Vitrolife, Gothenburg, Sweden). All embryos were cultured at 37°C under a gas phase of 5% O_2, 6% CO_2 and 89% N_2 from days 0–7.

Embryo observation, blastocyst monitoring and vitrification

Fertilization was assessed 16–20 h after ICSI. Normally fertilized zygotes with two pronuclei were cultured until the blastocyst stage. Embryos were observed using EmbryoViewer software without removing the culture dish from the incubator. Embryo observations were carried out to confirm the presence of two pronuclei on day 1. According to the centre protocol for vitrification, embryonic development was closely monitored between days 5 and 7 by twice-daily checks until blastocyst vitrification. For blastocyst vitrification on days 5 or 6, blastocysts were required to attain an inner diameter of over 160 μm and Gardner's grading CC or above (*Ueno et al., 2020*). Embryos that fulfilled these blastocyst vitrification criteria were included in the study and were defined as utilized blastocysts. The selected embryos were vitrified immediately according to the Cryotop method (*Mori et al., 2015*). Before starting the vitrification procedure, blastocysts were collapsed using a high-osmolality solution, according to a previous study (*Mukaida et al., 2006*). If the developing embryo did not fulfil the desired criteria for day 5 and 6, it was cultured until day 7. For blastocyst vitrification on day 7, blastocysts were required to attain an inner diameter wider than 180 μm and Gardner's grading CC or above (*Ueno et al., 2020*). If an embryo did not conform to these criteria by day 7, it was discarded. Measurements of the blastocyst inner diameter were taken using EmbryoViewer software.

Blastocyst grading

All transferred blastocysts were graded using three blastocyst assessment methods: iDA-V2, iDA-V1 and Gardner grading (*Gardner et al., 2000*). For each iDAScore version, the scores were grouped into quartiles (iDA-V2: 1.0–3.0, 3.1–5.5, 5.6–7.5, 7.6–9.9; iDA-V1: 1.0–7.3, 7.4–8.8, 8.9–9.3, 9.4–9.9). Well-trained embryologists carried out morphological grading for Gardner grading. The kappa statistic for blastocyst morphology was 0.6 in the inner cell mass (ICM) and 0.7 in the trophectoderm. The degree of expansion in all blastocysts was grade 4, owing to the cryopreservation strategy. The Gardner grading was, therefore, stratified into four groups (4AA, 4AB + 4BA, 4BB and others, including AC, CA, BC, CB, CC).

Post-warming embryo culture

Blastocysts for warming were selected based on in-house grading carried out before vitrification. After warming, assisted hatching was carried out by laser (Saturn 5) (Origio, Måløv, Denmark) for complete zona removal, according to our previous study (*Ueno et al., 2016*). Surviving blastocysts were cultured for 30 min to 2 h until blastocoel re-expansion was confirmed; blastocysts with the same or increased blastocoel size relative to the size before vitrification were transferred. Finally, degenerated blastocysts were discarded.

Single vitrified–warmed blastocyst transfer procedure

The endometrial preparation method was determined after the patient consulted during the SVBT cycles. In principle, SVBT was carried out after the confirmation of spontaneous ovulation in the natural cycle (*Kato et al., 2012*). Alternatively, hormone replacement transfer (HRT) cycles were selected in cases of severe ovulatory failure, ovarian insufficiency, luteal insufficiency or when SVBT in the ovulatory cycle was deemed inappropriate (*Takeshima et al., 2022*).

In the natural cycle, the ovulation date was estimated from the endometrial thickness, follicle size and hormonal patterns or determined by buserelin triggering as below. The timing of SVBT in the natural cycle was classified into two patterns according to the timing of ovulation: day 4.6 (ovulation 4.5), or day 5 (ovulation 5.0) after ovulation. When the serum LH level did not increase on the day of the trigger, SVBT was scheduled on day 7 after the trigger (ovulation 5.0). When the LH level slightly increased on the day of the trigger, SVBT was scheduled on day 6 after the trigger (ovulation 5.0).

The HRT cycle consists of the administration of exogenous oestradiol and progesterone. Oestradiol administration (Estrana) (Hisamitsu Pharmaceutical, Saga, Japan; Julina, Bayer Yakuhin, Tokyo, Japan) was initiated on the second day of menstruation. Dydrogesterone administration was initiated after confirmation that the endometrial thickness was greater than 7 mm and the serum oestradiol level reached 200 pg/ml. Progesterone was administered intravaginally 3 days after starting dydrogesterone administration. Single vitrified–warmed blastocyst transfers were carried out 4 days after starting progesterone administration.

Single vitrified–warmed blastocyst transfers were carried out as previously described (*Kato et al., 2012*; *Takeshima et al., 2022*). In addition, embryo transfer was carried out under vaginal ultrasound guidance using a specially designed soft silicone inner catheter (Kitazato ET catheter) (Kitazato Corporation, Shizuoka, Japan) by placing the blastocyst in a minimal volume in the upper part of the uterine cavity.

Luteal support was provided depending on the blood progesterone level on the day of embryo transfer. Patients with a blood progesterone level above 12 ng/ml on the day of embryo transfer were administered dydrogesterone (30 mg/day orally) (Daiichi Sankyo, Tokyo, Japan). Single vitrified–warmed blastocyst transfers were not carried out in patients with blood progesterone levels less than 8 ng/ml. Patients with progesterone levels in the range of 8–12 ng/ml were administered progesterone intravaginally (Luteum Vaginal Suppository) (ASKA Pharmaceutical Co., Ltd., Tokyo, Japan) until 8 weeks of pregnancy.

For clinical outcomes, the following definitions were used: biochemical pregnancy loss (serum HCG level over 20 IU/ml [*Ueno et al., 2014*] but no gestational sac); implantation or clinical pregnancy (a confirmed gestational sac at 6–7 weeks of pregnancy); ongoing pregnancy (mentioned above); first-trimester miscarriage (gestational sac confirmed, but no fetal heartbeat).

Statistical analysis

Ordinal variables were analysed using the Cochran-Armitage test for trends, and the Mann–Whitney U test was used to compare continuous variables. Multivariate

TABLE 1 PATIENT CHARACTERISTICS ACCORDING TO MATERNAL AGE GROUP

Characteristics	<35 years	35–37 years	38–40 years	41–42 years	≥43	All
Cycle, *n*	757	821	1007	715	660	3960
Maternal age, mean, ± SD	31.7 ± 2.1	36.0 ± 0.8	39.1 ± 0.8	41.5 ± 0.5	44.2 ± 1.2	38.3 ± 4.3
Paternal age, mean, ± SD	34.7 ± 4.4[a]	38.4 ± 4.2[b]	41.1 ± 4.8[c]	43.1 ± 4.8[d]	45.4 ± 5.3[e]	40.4 ± 5.9
Previous egg retrievals, *n*, mean, ± SEM	1.4 ± 0.1[a]	1.7 ± 0.1[b]	2.3 ± 0.1[c]	2.7 ± 0.1[d]	5.6 ± 0.1[e]	2.6 ± 0.1
Previous embryo transfers, *n*, mean, ± SE	0.8 ± 0.1[a]	1.2 ± 0.1[b]	1.5 ± 0.1[c]	1.7 ± 0.1[c]	2.7 ± 0.1[d]	1.5 ± 0.03
Cause of infertility						
Male factor	31.7 (240/757)	37.3 (306/821)	37.5 (378/1007)	37.2 (266/715)	39.2 (259/660)	36.6 (1449/3960)
Tubal factor	3.7 (28/757)	3.5 (29/821)	3.5 (35/1007)	2.4 (17/715)	2.0 (13/660)	3.0 (122/3960)
Endometriosis	4.2 (32/757)	4.8 (39/821)	3.1 (31/1007)	2.9 (21/715)	3.2 (21/660)	3.6 (144/3960)
Ovulation disorders	7.3 (55/757)	4.1 (34/821)	3.8 (38/1007)	2.5 (18/715)	1.1 (7/660)	3.8 (152/3960)
Uterine or cervical factor	2.9 (22/757)	4.6 (38/821)	4.8 (48/1007)	5.3 (38/715)	4.7 (31/660)	4.5 (177/3960)
Combination	20.6 (156/757)	16.6 (136/821)	16.6 (167/1007)	16.5 (118/715)	16.1 (106/660)	17.2 (683/3960)
Unknown	29.5 (223/757)	29.0 (238/821)	30.7 (309/1007)	33.0 (236/715)	33.8 (223/660)	31.0 (1229/3960)
Other	0.1 (1/757)	0.1 (1/821)	0.1 (1/1007)	0.1 (1/715)	0.0 (0/660)	0.1 (4/3960)

Age groups were stratified according to Society for Assisted Reproductive Technology age groups. Values are expressed as % (n/N), unless otherwise stated. Different superscript letters (a, b, c, d, e) within a row indicate significant differences among the maternal age group (paternal age, $P < 0.0001$), number of previous eggs retrievals ($P < 0.0001$ except for the group aged 38–40 years compared with the group aged 41–42 years ($P = 0.0016$) and group aged younger than 35 years compared with the group aged 35–37 years ($P = 0.009$)). The number of previous embryo transfers differed significantly ($P < 0.0001$ except for the group aged 35–37 years compared with the group aged 38–40 years ($P = 0.0032$). The Mann–Whitney U test was used to compare continuous variables.

logistic regressions were used to analyse the relationship between clinical outcome (ongoing pregnancy-positive or negative) after SVBT and patients, embryo characteristics, or both. The discrimination performance of the iDA-V2, iDA-V1, and Gardner grading for predicting ongoing pregnancy, was calculated using ROC analysis. The AUCs were compared using a paired two-tailed DeLong test. An AUC of 0.5 is equivalent to random classification, and 1.0 is equivalent to 100% correct classification.

The JMP software (version 10.0; SAS Institute, Cary, NC, USA) and R (version 3.6.1, 2019-07-05) were used for all statistical analyses.

RESULTS

The participant characteristics for each maternal age group are presented in TABLE 1. Paternal age (all comparisons $P < 0.0001$), number of previous eggs retrievals ($P < 0.0001$ for all comparisons except the group aged 38–40 years compared with the group aged 41–42 years [$P = 0.0016$] and the group aged <35 years compared with the group aged 35–37 years [$P = 0.009$]) differed significantly between the maternal age groups. The number of previous embryo transfers differed significantly ($P < 0.0001$ for all comparisons except for the group aged 35–37 years compared with the group aged 38–40 years [$P = 0.0032$]) between the maternal age groups, apart from the group aged 38–40 years compared with the group aged 41–42 years ($P = 0.2428$). The analyses in this study, however, were conducted for each maternal age group; therefore, these factors did not introduce bias. The distribution of scores and grades in the iDA-V2, iDA-V1 and Gardening grading are presented in TABLE 2. For iDA-V2, the group aged 35–37 years showed no significant difference in the scores with the group aged younger than 35 years ($P = 0.1505$). A significant difference was found in the scores among the other groups ($P < 0.0001$ for all comparisons). For iDA-V1, however, no significant differences were found among groups aged younger than 35 years (compared with the group aged 35–37 years, $P = 0.9998$ and with the group aged 38–40 years, $P = 0.1150$), the group aged 35–37 years (compared with the group aged 38–40 years, $P = 0.0630$) and the group aged 38–40 years. The others demonstrated a significant difference in their scores ($P < 0.0001$ except for the group aged 38–40 years compared with the group aged 41–42 years ($P = 0.018$), the group aged 41–42 years and the group aged 43 years and over ($P = 0.0126$). Box plots for the distributions of iDA-V2 and iDA-V1 for ongoing pregnancy-positive and negative blastocysts within each age group are presented in FIGURE 1. The figure shows that, for all age groups, the scores of iDA-V2 were distributed over a wider range whereas those of iDA-V1 were distributed with a narrower range.

Clinical outcomes after SVBT, stratified by the iDA-V2 group, are presented in TABLE 3. For beta-HCG positive ($P < 0.0001$), clinical pregnancy ($P < 0.0001$) and ongoing pregnancy ($P < 0.0001$), a statistically highly significant score-dependent decline was observed with decreasing score across the quartiles according to Cochran-Armitage trend test. For biochemical pregnancy loss ($P = 0.0004$) and first-trimester miscarriage ($P < 0.0001$), a statistically highly significant score-dependent increase was observed with decreasing score across the quartiles according to Cochran-Armitage trend test.

Ongoing pregnancy rates for iDAScore v2.0 and iDAScore v1.0

Before comparing the AUC, the trend between ongoing pregnancy rates and iDA-V2 (FIGURE 2A) or iDA-V1 (FIGURE 2B) scores was analysed. iDA-V2 and iDA-V1 were presented as quartiles, not

TABLE 2 BLASTOCYST SCORING GRADE IN EACH MODEL ACCORDING TO MATERNAL AGE GROUP

		<35 years	35–37 years	38–40 years	41–42 years	≥43 years	All
Blastocyst, *n*		757	821	1007	715	660	3960
iDA-V2	Mean ± SD	5.87 ± 2.40[a]	5.74 ± 2.34[a,b]	5.44 ± 2.41[b]	4.92 ± 2.35[c]	4.43 ± 2.28[d]	5.32 ± 2.41
	Maximum	9.4	9.7	9.7	9.2	9.5	9.7
	75%	8.0	7.8	7.7	7.1	6.3	7.6
	Median	6.5	6.2	5.8	5.0	4.3	5.6
	25%	3.8	3.9	3.3	2.6	2.3	3.1
	Minimum	1.0	1.0	1.0	1.0	1.1	1.0
iDA-V1	Mean ± SD	8.32 ± 1.52[a]	8.34 ± 1.45[a]	8.14 ± 1.54[a]	7.91 ± 1.55[b]	7.65 ± 1.65[c]	8.09 ± 1.56
	Maximum	9.9	9.9	9.9	9.8	9.7	9.9
	75%	9.4	9.3	9.3	9.2	9.0	9.3
	Median	9.0	9.0	8.8	8.5	8.1	8.8
	25%	8.0	7.9	7.5	6.8	6.6	7.3
	Minimum	2.5	2.1	1.9	2.2	2.3	1.9
Gardner grading	4AA[e]	38.2	38.0	33.4	27.3	20.0	31.9
	4AB, 4BA	20.6	19.6	18.9	20.3	18.2	19.5
	4BB	10.3	11.1	11.7	11.6	10.9	11.1
	Others[e]	30.9	31.3	36.1	40.8	50.9	37.4

Different superscript letters (a, b, c, d) within a row indicate significant differences. In the iDAScore version 2 (iDA-V2) group, all significant differences are $P < 0.0001$. For iDAScore version 1 (iDA-V1) $P < 0.0001$ except for the group aged 38–40 years compared with the group aged 41–42 years ($P = 0.0186$) and the group aged 41–42 years compared with the group aged 43 years and over ($P = 0.0126$). Mann–Whitney U test was used to compare continuous variables. 'Others' in Gardner grading indicates 4AC, 4CA, 4BC, 4CB and 4CC.

[e] $P < 0.0001$ by Cochran–Armitage trend test.

continuous variables (iDA-V2: excellent, 9.9–7.6; good, 7.5–5.6; fair, 5.5–3.1; poor, 3.0–1.0; iDA-V1: excellent, 9.9–9.4; good, 9.3–8.9; fair, 8.8–7.4; poor, 7.3–1.0). The ongoing pregnancy rates were calculated by dividing the score equally (interval: 1.0). iDA-V2 showed a more pronounced linear increase in ongoing pregnancy rates with increasing scores than iDA-V1 (FIGURE 3).

Comparison of the area under the curve for ongoing pregnancy between iDAScore v2.0 and iDAScore v1.0 or Gardner grading

The AUCs were compared between iDA-V2 and iDA-V1 or Gardener grading for each maternal age group (TABLE 4). Overall, iDA-V2 was significantly higher than iDA-V1 and Gardener grading ($P < 0.0001$). In the AUCs of the group aged younger than 35 years ($n = 757$) for ongoing pregnancy prediction, the AUC of iDAV2 was significantly higher than that of Gardener grading (0.733 versus 0.694, $P = 0.015$). For the group aged 35–37 years ($n = 821$), the AUC of iDA-V2 was significantly higher than that of iDA-V1 (0.712 versus 0.696, $P = 0.035$). In the AUCs of the group aged 38–40 years ($n = 1007$), no significant differences were observed. In the AUCs of the group aged 41–42 years ($n = 715$), the AUC of iDA-V2 was significantly higher from that of Gardener grading (0.745 versus 0.696, $P = 0.007$). In the group aged over 42 years ($n = 660$), no significant difference was found among the groups. The ROC curves for iDA-V2 and iDA-V1 for each maternal age group are presented in the **Supplementary Figure**.

Comparison of area under the curve for ongoing pregnancy between iDAScore v2.0 and iDAScore v1.0 within different blastocyst sub-groups

The results of the comparison of AUC between iDA-V2 and iDA-V1 for different blastocyst sub-groups are presented in TABLE 5. The AUC for iDA-V2 was significantly higher than iDA-V1 in all sub-groups for the day of vitrification (day 5 and day 6; both $P < 0.0001$), ICM ($P < 0.0001$, $P = 0.032$ and $P = 0.008$ for grades A, B and C, respectively) and blastocyst inner diameters ($P = 0.032$, $P = 0.006$, $P = 0.031$ and $P = 0.009$ for > 189 μm, 189–180 μm, 179–170 μm, and <170 μm, respectively). The differences in AUC were also significantly higher for iDA-V2 compared with iDA-V1 for trophectoderm sub-groups A ($P = 0.025$) and B ($P = 0.002$) but not for grade C ($P = 0.107$).

Contribution of the embryo score to ongoing pregnancy in iDAScore v2.0 and iDAScore v1.0

The results of the multivariate logistic regression analysis for the comparison of the contribution of each iDAScore version to ongoing pregnancy are presented in TABLE 6. Multivariable logistic regression analysis, which included maternal and paternal age, number of previous IVF treatments, number of previous embryo transfers, number of retrieved oocytes, endometrial thickness, insemination methods (conventional insemination versus ICSI), endometrial preparation (ovulation induction embryo transfer versus HRT), luteal phase support protocol (dydrogesterone to luteum), ICM grade, trophectoderm grade, day of vitrification, blastocyst diameters and iDA-V2 or iDA-V1 were used as confounding factors to compare the contribution of the embryo score with pregnancy in each iDAScore version. A significant correlation was found for pregnancy in iDA-V2 (adjusted odds ratio [aOR] 1.11, 95% CI 1.04 to 1.17, $P = 0.0011$), maternal age (aOR 0.84, 95% CI 0.82 to 0.86, $P < 0.0001$), ICM grade (Grade A versus C: aOR 0.62, 95% CI 0.46 to 0.84, $P = 0.002$), trophectoderm grade (A versus B: aOR 0.66, 95% CI 0.54

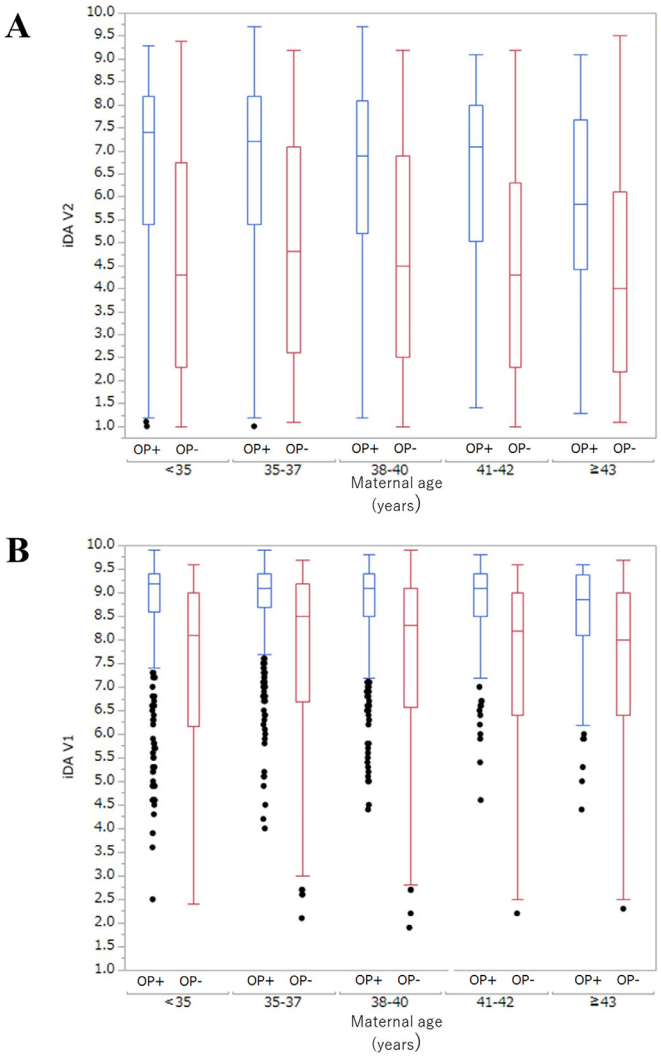


FIGURE 1 Box plots for the distribution of the iDAScore v1.0 (IDA-V1) and iDAScore v2.0 (IDA-V2) for ongoing pregnancy-positive (OP+) and ongoing pregnancy-negative (OP−) blastocysts within each age group. Box plots show median, interquartile range, maximum and minimum, and outliers.

to 0.80, $P < 0.0001$; A to C: aOR 0.43, 95% CI 0.32 to 0.56, $P < 0.0001$), day of vitrification (day 5 versus day 6: aOR 0.46, 95% CI 0.35 to 0.62, $P < 0.0001$; day 5 versus day 7: aOR 0.07, 95% CI 0.01 to 0.52, $P = 0.0100$) and blastocyst inner diameter (aOR 1.01, 95% CI 1.01 to 1.02, $P < 0.0001$). In version 1, a significant correlation was found for pregnancy in maternal age (aOR 0.84, 95% CI 0.82 to 0.86, $P < 0.0001$), insemination methods (conventional IVF versus ICSI: aOR 1.23, 95% CI 1.04 to 1.46, $P = 0.013$), ICM grade (Grade A versus C: aOR 0.62, 95% CI 0.45 to 0.83, $P = 0.0019$), trophectoderm grade (A versus B: aOR 0.61, 95% CI 0.51 to 0.74, $P < 0.0001$; A versus C: aOR 0.37, 95% CI 0.28 to 0.49, $P < 0.0001$), day of vitrification (day 5 versus day 6: aOR 0.40, 95% CI 0.28 to 0.55, $P < 0.0001$; day 5 versus day 7: aOR 0.06, 95% CI 0.01 to 0.45, $P = 0.0064$) and blastocyst inner diameter (aOR 1.01, 95% CI 1.00 to 1.02, $P < 0.0001$). No significant correlation, however, was found between iDA-V1 and ongoing pregnancy (aOR 1.06, 95% CI 0.95 to 1.19, $P = 0.292$), although, in univariate logistic regression analysis, iDA-V1 has a significant correlation with ongoing pregnancy (OR 1.89, 95% CI 1.77 to 2.01, $P < 0.0001$).

DISCUSSION

Deep learning models have been used in medical fields, including IVF treatment. For pregnancy prediction in IVF, our previous study suggested that the iDAScore model has a higher performance than morphokinetic and morphological grading models (*Ueno et al., 2021*). The AUC for

TABLE 3 CLINICAL OUTCOMES AFTER SINGLE VITRIFIED–WARMED BLASTOCYST TRANSFER STRATIFIED IDASCORE V2.0 GROUPS

	iDAScore v2.0 score quartiles				
Outcomes	**9.9–7.6**	**7.5–5.6**	**5.5–3.1**	**3.0–1.0**	***P*-value**
Transferred blastocysts, *n*	1007	974	1021	958	NA
Maternal age at egg retrieval, mean, ±SD	37.0 ± 4.1[a]	38.0 ± 4.2[b]	39.0 ± 4.1[c]	39.2 ± 4.4[c]	NA
β-HCG positive/SVBT, % (*n*/N)	74.7 (752/1007)	59.1 (576/974)	41.2 (421/1021)	20.6 (197/958)	<0.0001
Biochemical pregnancy loss/β-HCG positive, % (n/N)	6.8 (51/752)	10.9 (63/576)	9.7 (41/421)	15.7 (31/197)	0.0004
Clinical pregnancy/SVBT, % (*n*/N)	69.6 (701/1007)	52.7 (513/974)	37.2 (380/1021)	17.3 (166/958)	<0.0001
First-trimester miscarriage/clinical pregnancy, % (*n*/N)	7.7 (54/701)	8.8 (45/513)	14.5 (55/380)	21.1 (35/166)	<0 .0001
Ongoing pregnancy/SVBT, % (*n*/N)	64.3 (647/1007)	48.0 (468/974)	31.8 (325/1021)	13.7 (131/958)	<0.0001

iDAScore groups were stratified by quartiles. Different superscript letters (a, b, c) within a row indicate significant differences. For maternal age $P < 0.0001$ except for the group 5.5–3.1 versus the group 3.0–1.0 group ($P = 0.0891$). Statistical analysis used Cochran–Armitage trend test and for comparing continuous variables, Mann–Whitney U test was used. β-HCG, beta HCG; NA, not applicable; SVBT, single vitrified–warmed blastocyst transfer.

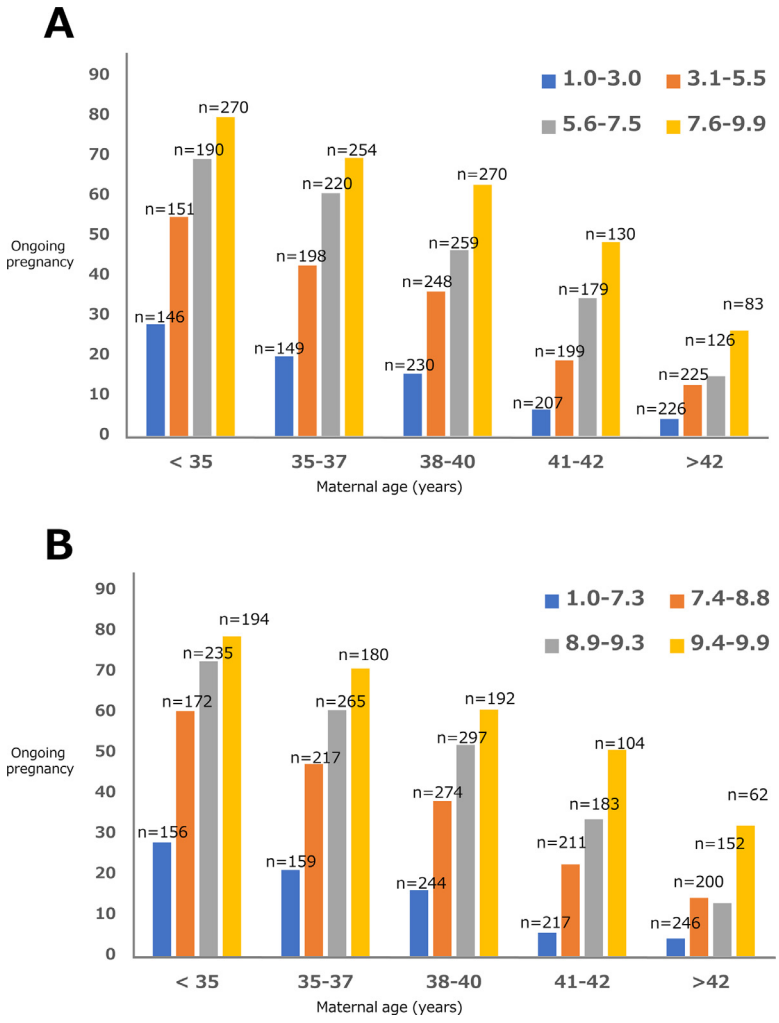


FIGURE 2 (A) Ongoing pregnancy rates on each iDAScore v2.0 (iDA-V2) score group stratified by Society for Assisted Reproductive Technology (SART) maternal age criteria; (B) ongoing pregnancy rates for each iDAScore v1.0 (iDA-V1) score group stratified by SART maternal age criteria. Ongoing pregnancy rates were significantly different ($P < 0.0001$) between the score groups within each age group for each model according to Cochran-Armitage trend test.

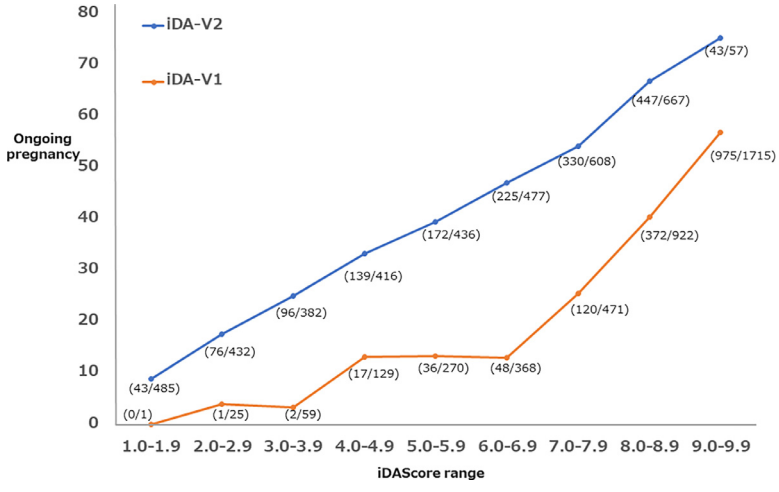


FIGURE 3 Ongoing pregnancy rates according to iDAScore v1.0 (IDA-V1) and iDAScore v2.0 (IDA-V2) divided into sub-groups with equal size (interval: 1.0).

pregnancy prediction using the first commercial iDAScore model, however, was 0.700, suggesting that model performance could potentially be improved even further.

In general, the performance of deep learning models can be improved by either changing the training data, the model architecture, training methods and evaluations. For the update of iDAScore from iDA-V1 to iDA-V2, the same deep-learning architecture was used, but with a 15% larger training dataset for day 5 and slightly modified training methods (*Lassen et al., 2023*). The slightly modified training methods include new data augmentation strategies, new hyper-parameters, knowledge distillation, calibration and other smaller adjustments (*Lassen et al., 2023*). In agreement with the current results *Lassen et al. (2023)* found that iDA-V2 also had a significantly higher AUC than iDA-V1 for all age groups. They reported an increase in AUC of 0.022 from 0.672 to 0.694, which is similar to the observed increase from 0.720 to 0.736 in the present study. It should be noted that the actual AUC value should not be compared owing to differences in clinical practices and patient cohorts between the two data sets (*Kragh and Karstoft, 2021*).

Although the AUC within all age sub-groups seemed to be higher for iDA-V2 compared with iDA-V1, the difference was only statistically significant for the group aged 35–37 years, as well as for the overall data. To our knowledge, this is the first study indicating that increasing training data for deep-learning pregnancy prediction models improved the performance of pregnancy prediction. Interestingly the difference in AUC between the different age-groups within the two models shows the same pattern except for the group aged 38–40 years compared with the group aged 35–37 years. This indicates that the differences in AUC between the age groups are not as much attributed to internal model bias but rather that the cohort of transferred blastocysts within the age-subgroups are fundamentally different. This is further supported by the finding that older patients have a lower fraction of good-quality blastocysts (TABLE 2). For both models, the discrimination performance was lowest for the poorest quality blastocysts, corresponding with the group aged younger than 42 years, whereas the highest discrimination was seen in the group aged 41–42 years, which showed

TABLE 4 COMPARISON OF THE AREA UNDER THE CURVE FOR ONGOING PREGNANCY BETWEEN THREE EMBRYO SCORING AND GRADING MODELS WITHIN EACH MATERNAL AGE GROUP

Age group, years	AUC of iDA-V2 (95% CI)	AUC of iDA-V1 (95% CI)	AUC of GG (95% CI)	*P*-value iDA-V2 versus iDA-V1	iDA-V2 versus GG
<35	0.733 (0.695 to 0.771)	0.718 (0.680 to 0.756)	0.694 (0.657 to 0.731)	0.059	0.015
35–37	0.712 (0.678 to 0.747)	0.696 (0.660 to 0.732)	0.695 (0.660 to 0.729)	0.035	0.242
38–40	0.706 (0.674 to 0.738)	0.698 (0.666 to 0.730)	0.700 (0.669 to 0.731)	0.185	0.667
41–42	0.745 (0.705 to 0.786)	0.734 (0.693 to 0.774)	0.696 (0.655 to 0.737)	0.174	0.007
>42	0.698 (0.640 to 0.756)	0.685 (0.625 to 0.744)	0.682 (0.624 to 0.741)	0.250	0.538
Total	0.736 (0.720 to 0.751)	0.720 (0.705 to 0.736)	0.702 (0.686 to 0.718)	< 0.0001	< 0.0001

Areas under the curve were compared using a paired two-tailed DeLong test.

AUC, area under the curve; GG, Gardener grading; iDA-V1, iDAScore version 1; iDA-V2, iDAScore version 2.

poorer quality blastocysts than the younger age groups, indicating that discrimination was not highest for age groups with the highest proportion of good-quality blastocysts. Both models seem to be less able to sort the highest-quality blastocysts within some age groups, suggesting that, for some blastocysts, the score does not reflect implantation potential. Therefore, we hypothesize that the observed performance differences between age subgroups are related to differences in the quality of the transferred blastocysts.

For the blastocyst quality sub-groups, the discrimination performance of iDA-V2 was always higher than iDA-V1. The difference was significant for all sub-groups except for trophectoderm grade C. Therefore, because of the larger training data, the updated version has been able to learn the subtler differences between similarly looking blastocysts. This is also evident from the multivariable logistic regression analysis in which patient factors and blastocyst quality are included. Here, iDA-V2 had a significantly higher adjusted odds ratio, whereas the adjusted odds ratio for iDA-V1 was not significant. This result indicated that iDA-V2 is an independent factor for prediction of pregnancy after SVBT. The non-significance of iDA-V1 indicates that there is no added benefit of evaluating blastocysts with both, static observation parameters including Gardner grades, blastocyst diameters, day of vitrification and iDA-V1. It should be noted, however, that iDA-V1 is significant in a univariate logistic regression and that the overall AUC is higher for iDA-V1 than Gardner grades when evaluated on the whole dataset. In addition, iDAScore models do not require any manual annotations that are prone to both subjectivity and time-consumption.

TABLE 5 RESULTS OF COMPARISON OF AREAS UNDER THE CURVE FOR ONGOING PREGNANCY BETWEEN IDASCORE VERSION 2 AND IDASCORE VERSION 1 WITHIN BLASTOCYST SUB-GROUPS

Characteristics	Parameters, *n*	iDA-V2	iDA-V1	*P*-value
Day of vitrification[a]	Day 5 (2777)	0.652	0.627	<0.0001
	Day 6 (1144)	0.691	0.643	<0.0001
Inner cell mass	A (1901)	0.653	0.629	<0.0001
	B (1227)	0.673	0.657	0.032
	C (832)	0.726	0.692	0.008
Trophectoderm	A (1521)	0.615	0.595	0.025
	B (1066)	0.592	0.561	0.002
	C (1373)	0.700	0.685	0.107
Diameter group, μm	>189 (1250)	0.742	0.728	0.032
(quartile)	189–180 (779)	0.701	0.679	0.006
	179–170 (955)	0.720	0.627	0.031
	<170 (976)	0.730	0.643	0.009

[a] Day-7 vitrification was not included owing to a low number of blastocysts transferred (n = 39).

Areas under the curve were compared using a paired two-tailed DeLong test.

iDA-V1, iDAScore version 1; iDA-V2, iDAScore version 2.

Both iDA-V2 and iDA-V1 use a score scale of 1.0–9.9 in which blastocysts with a grade of 9.9 have the highest likelihood of pregnancy. Overall, the ongoing pregnancy rates in the sub-group 9.0–9.9 were 75% and 57% for iDA-V2 and iDA-V1, respectively (FIGURE 3). In iDA-V1, no Platt scaling was carried out, and iDA-V2 was calibrated to the observed ongoing pregnancy rates; therefore, the scaling of scores has changed between the two models (*Lassen et al., 2023*). Furthermore, iDA-V2 showed a more pronounced linear increase in ongoing pregnancy rates with increasing scores than iDA-V1 (FIGURE 3). This means that the updated calibration procedure ensured a linear correlation between scores and the likelihood of ongoing pregnancy.

The iDA-V2 score was related to clinical outcomes, including positive beta-HCG, biochemical pregnancy loss, clinical pregnancy, first-trimester miscarriage and ongoing pregnancy. As our previous study suggested that biochemical pregnancy loss rates were not related with iDA-V1 (*Ueno et al., 2022*), the correlation with biochemical pregnancy loss rate shown in the present study represents an improvement. This result suggests that using iDA-V2 allows for improved clinical outcome prediction compared with iDA-V1.

In the present study, blastocyst scoring was only carried out before vitrification. In our clinic, blastocyst quality is not re-scored after warming. Therefore, we have no data about changes in the score before vitrification and after warming. Vitrification

TABLE 6 RESULTS OF MULTIVARIABLE LOGISTIC REGRESSION ANALYSIS FOR ONGOING PREGNANCY AFTER SINGLE VITRIFIED–WARMED BLASTOCYST TRANSFER IN IDASCORE VERSION 2 AND IDASCORE VERSION 1

	iDA-V2			iDA-V1		
Characteristics	**Adjusted odds ratio**	**95% CI**	***P*-value**	**Adjusted odds ratio**	**95% CI**	***P*-value**
Maternal age	0.84	0.82 to 0.86	<0.0001	0.84	0.82 to 0.86	<0.0001
Paternal age	1.00	0.99 to 1.02	0.665	1.00	0.99 to 1.02	0.616
Number of IVF treatments	0.99	0.95 to 1.03	0.481	0.98	0.95 to 1.02	0.410
Number of embryo transfers	1.01	0.95 to 1.07	0.751	1.01	0.96 to 1.07	0.773
Number of retrieved oocytes	1.03	0.98 to 1.08	0.287	1.03	0.98 to 1.08	0.253
Endometrial thickness	1.02	0.98-1.06	0.300	1.02	0.98 to 1.08	0.300
Insemination methods (C-IVF versus ICSI)	1.18	0.99 to 1.40	0.051	1.23	1.04 to 1.46	0.013
Endometrial preparation (OVT versus HRT)	0.73	0.53 to 1.01	0.057	1.37	0.99 to 1.89	0.058
Luteal phase support protocol (dydrogesterone to luteum)	0.88	0.70 to 1.11	0.297	0.88	0.70 to 1.11	0.287
ICM grade						
A	Reference			Reference		
B	1.02	0.84 to 1.23	0.853	0.99	0.83 to 1.21	0.940
C	0.62	0.46 to 0.84	0.002	0.62	0.45 to 0.83	0.0019
TE grade						
A	Reference			Reference		
B	0.66	0.54 to 0.80	<0.0001	0.61	0.51 to 0.74	<0.0001
C	0.43	0.32-0.56	<0.0001	0.37	0.28 to 0.49	<0.0001
Vitrification day						
5	Reference			Reference		
6	0.46	0.35 to 0.62	<0.0001	0.40	0.28 to 0.55	<0.0001
7	0.07	0.01 to 0.52	0.0100	0.06	0.01 to 0.45	0.0064
Blastocyst diameter	1.01	1.01 to 1.02	<0.0001	1.01	1.00 to 1.02	<0.0001
iDAScore	1.11	1.04 to 1.17	0.0011	1.06	0.95 to 1.19	0.292

aOR, adjusted odds ratio; C-IVF, conventional IVF; HRT, hormone replacement embryo transfer;

ICM, inner cell mass; ICSI, intracytoplasmic sperm injection; iDA-V1, iDAScore version 1;

iDA-V2, iDAScore version 2; OVT, ovulation induction embryo transfer; TE, trophectoderm.

and warming procedures can potentially damage blastocyst cells. Therefore, the inclusion of either a manual or an evaluation of the post-warming blastocyst based on artificial intelligence could potentially improve the prediction of the likelihood of pregnancy. A major limitation of our study is that it was based on minimal stimulation and natural IVF treatment cycle; therefore, this cohort contained few situations in which elective blastocyst transfer occurred. Further studies using randomized controlled trials will be crucial to analyse the efficacy of selection in elective blastocyst transfer situations. Furthermore, this study only compared two deep-learning blastocyst assessment models that did not only differ in sample size.

In conclusion, the present study suggests that the performance of deep learning models for pregnancy prediction will be improved by increasing the size of the training data. In addition, an updated calibration procedure ensured a linear correlation between scores and the likelihood of pregnancy. Therefore, in a deep-learning model for pregnancy prediction, continuously collecting training data is an important step for improving the performance of updated artificial intelligence models.

DATA AVAILABILITY

Data will be made available on request.

ACKNOWLEDGEMENTS

The authors wish to thank Markus Montag, PhD (ilabcomm GmbH), for his help editing the initial manuscript draft. The authors also thank the clinical embryologists for their efforts in the laboratory and for their assistance with this study.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.rbmo.2023.103308.

REFERENCES

Gardner, DK, Lane, M, Stevens, J, Schlenker, T, Schoolcraft, WB, 2000. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. Fertil. Steril. 73, 1155–1158.

Glatstein, I, Chavez-Badiola, A, Curchoe, CL, 2023. New frontiers in embryo selection. J Assist Reprod Genet In press.

Kato, K, Ueno, S, Berntsen, J, Kragh, MF, Okimura, T, Kuroda, T, 2023. Does embryo categorization by existing artificial intelligence, morphokinetic or morphological embryo selection models correlate with blastocyst euploidy rates? Reprod BioMed Online 46 (2), 274–281.

Kato, K, Takehara, Y, Segawa, T, Kawachiya, S, Okuno, T, Kobayashi, T, Bodri, D, Kato, O, 2012. Minimal ovarian stimulation combined with elective single embryo transfer policy: age-specific results of a large, single-center,Japanese cohort. Reprod. Biol. Endocrinol. 10, 35.

Kragh, MF, Karstoft, H, 2021. Embryo selection with artificial intelligence: how to evaluate and compare methods? J. Assist. Reprod. Genet 38, 1675–1689.

Kragh, MF, Rimestad, J, Lassen, JT, Berntsen, J, Karstoft, H, 2022. Predicting Embryo Viability Based on Self-Supervised Alignment of Time-Lapse Videos. IEEE Trans Med Imaging 41, 465–475.

Lassen, JT, Kragh, MF, Rimestad, J, Nygard, M, Berntsen, J, 2023. Development and validation of deep learning based embryo selection across multiple days of transfer. Sci. Rep. 13, 4235.

Moons, KG, Altman, DG, Reitsma, JB, Ioannidis, JP, Macaskill, P, Steyerberg, EW, Vickers, AJ, Ransohoff, DF, Collins, GS, 2015. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine. Ann. Intern. Med. 162, W1–W73.

Mori, C, Yabuuchi, A, Ezoe, K, Murata, N, Takayama, Y, Okimura, T, Uchiyama, K, Takakura, K, Abe, H, Wada, K, et al., 2015. Hydroxypropyl cellulose as an option for supplementation of cryoprotectant solutions for embryo vitrification in human assisted reproductive technologies. Reprod. Biomed. Online 30, 613–621.

Mukaida, T, Oka, C, Goto, T, Takahashi, K, 2006. Artificial shrinkage of blastocoeles using either a micro-needle or a laser pulse prior to the cooling steps of vitrification improves survival rate and pregnancy outcome of vitrified human blastocysts. Hum. Reprod. 21, 3246–3252.

Takeshima, K, Ezoe, K, Onogi, S, Kawasaki, N, Hayashi, H, Kuroda, T, Kato, K, 2022. Endometrial preparation and maternal and obstetrical outcomes after frozen blastocyst transfer. AJOG Glob. Rep. 2, 100081.

Teramoto, S, Kato, O, 2007. Minimal ovarian stimulation with clomiphene citrate a large-scale retrospective study. Reprod. Biomed. Online 15, 134–148.

Ueno, S, Ezoe, K, Abe, T, Yabuuchi, A, Uchiyama, K, Okuno, T, Kobayashi, T, Kato, K, 2014. Maternal age and initial beta-hCG levels predict pregnancy outcome after single vitrified-warmed blastocyst transfer. J. Assist. Reprod. Genet 31, 1175–1181.

Ueno, S, Uchiyama, K, Kuroda, T, Okimura, T, Yabuuchi, A, Kobayashi, T, Kato, K, 2020. Establishment of day 7 blastocyst freezing criteria using blastocyst diameter for single vitrified-warmed blastocyst transfer from live birth outcomes: a single-center, large cohort, retrospectively matched study. J. Assist. Reprod. Genet. 37, 2327–2335.

Ueno, S, Ezoe, K, Yabuuchi, A, Uchiyama, K, Okimura, T, Okuno, T, Kobayashi, T, Kato, K, 2016. Complete zona pellucida removal from vitrified-warmed human blastocysts facilitates earlier in-vitro attachment and outgrowth. Reprod. Biomed. Online 33, 140–148.

Ueno, S, Berntsen, J, Ito, M, Uchiyama, K, Okimura, T, Yabuuchi, A, Kato, K, 2021. Pregnancy prediction performance of an annotation-free embryo scoring system on the basis of deep learning after single vitrified-warmed blastocyst transfer: a single-center large cohort retrospective study. Fertil. Steril. 116, 1172–1180.

Ueno, S, Berntsen, J, Ito, M, Okimura, T, Kato, K, 2022. Correlation between an annotation-free embryo scoring system based on deep learning and live birth/neonatal outcomes after single vitrified-warmed blastocyst transfer: a single-centre, large-cohort retrospective study. J. Assist. Reprod. Genet. 39, 2089–2099.

Xue, Y, 2019. An overview of overfitting and its solutions. Journal of physics: Conference series. IOP Publishing, 022022.

Zaninovic N, Rosenwaks Z. Artificial intelligence in human in vitro fertilization and embryology.

Zhan, Q, Sierra, ET, Malmsten, J, Ye, Z, Rosenwaks, Z, Zaninovic, N, 2020. Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential. F. S. Rep. 1, 133–141.

Received 11 April 2023; received in revised form 20 June 2023; accepted 24 July 2023.