



Kent Academic Repository

Palmer, Giles A, Chavez-Badiola, Alejandro, Valencia-Murillo, Roberto, Harvey, Simon C., Mendizabal-Ruiz, Gerardo, Farías, Adolfo Flores-Saiffe, Paredes, Omar and Griffin, Darren K. (2025) *Can artificial intelligence guided feedback improve embryologists' selection of euploid embryos based on morphology alone?* Reproductive BioMedicine Online, 51 (4). ISSN 1472-6483.

Downloaded from

<https://kar.kent.ac.uk/111139/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.rbmo.2025.104990>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

ARTICLE

Can artificial intelligence guided feedback improve embryologists' selection of euploid embryos based on morphology alone?



BIOGRAPHY

Giles Palmer, an embryologist, holds expertise in laboratory, business and quality management. He studied genetics at Leeds University and achieved the first births in Greece using preimplantation genetic testing. He is an Honorary Lecturer at Bristol University, UK, Executive Director of the International IVF Initiative and Director of Communications for IVF2.0 Ltd.

Giles A. Palmer^{a,b,c}, Alejandro Chavez-Badiola^{b,d,e}, Roberto Valencia-Murillo^d, Simon C. Harvey^f, Gerardo Mendizabal-Ruiz^{d,g}, Adolfo Flores-Saiffe Farías^d, Omar Paredes^{b,h}, Darren K. Griffin^{i,j,k,*}

KEY MESSAGE

Embryologists and 'ERICA' identified aneuploid embryos better than random. ERICA outperformed human judgement, choosing euploid embryos on first attempt. No difference was found between embryologists in picking a better or worse embryo. No significant change occurred given the opportunity to change their mind after ERICA output.

ABSTRACT

Research question: Can embryologists reliably differentiate euploid from aneuploid embryos based on morphology alone, and can artificial intelligence (AI)-assisted selection, specifically using ERICA (Embryo Ranking Intelligent Classification Algorithm), improve embryo selection outcomes compared with embryologists alone?

Design: A training tool was developed and used in which 19 embryologists (comprising junior, intermediate and experienced practitioners) evaluated the ploidy status of embryo images. They were subsequently provided with rankings generated by ERICA

^a International IVF Initiative, New York, New York, USA

^b IVF2.0 Ltd, London, UK

^c Institute of Life, IASO Hospital Athens, Greece

^d Conceivable Life Sciences, New York, New York, USA

^e Hope IVF, Mexico City, Mexico

^f Faculty of Engineering and Science, University of Greenwich, UK

^g Computational Perception Laboratory, Department of Translational Bioengineering, University of Guadalajara, Mexico

^h Biodigital Innovation Laboratory, Department of Translational Bioengineering, CUCEI, Universidad of Guadalajara, Mexico

ⁱ School of Natural Sciences, University of Kent, Canterbury, Kent, CT27NJ, UK

^j Animal Genomics and Bioresource Research Unit (AGB Research Unit), Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand

^k UCL EGA Institute for Women's Health, Preimplantation Genetics Group, 86-96 Chenies Mews, London WC1E 6HX, UK

KEYWORDS

Artificial intelligence
Embryo selection
Aneuploidy detection
Embryologist decision-making

© 2025 The Author(s). Published by Elsevier Ltd on behalf of Reproductive Healthcare Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

*Corresponding author. E-mail address: d.k.griffin@kent.ac.uk (D.K. Griffin). <https://doi.org/10.1016/j.rbmo.2025.104990>
1472-6483/© 2025 The Author(s). Published by Elsevier Ltd on behalf of Reproductive Healthcare Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Declaration: Six authors are stakeholders in IVF technology companies. The companies involved are IVF2.0 Ltd, London, UK and Conceivable Life Sciences, New York, New York, USA. The patent for the software developed for the synthetic cycle creation process and embryo ranking application is owned by IVF2.0 Ltd. Giles Anthony Palmer is an employee of IVF2.0 Ltd, London, UK, a company specializing in developing AI software for assisted reproduction laboratories. Omar Paredes serves as the Head of Research at IVF2.0 Ltd, London, UK, a leader in AI-driven solutions for assisted reproduction laboratories. Alejandro Chavez-Badiola is a cofounder and shareholder of IVF2.0 Ltd, London, UK. He is also a cofounder and shareholder of Conceivable Life Sciences, New York, USA, a company that has pioneered the automation of clinical embryology processes through AI, computer imaging, and robotics. Roberto Valencia-Murillo, Gerardo Mendizabal-Ruiz, and Adolfo Flores-Saiffe Farías are former employees of IVF2.0 Ltd, where they were instrumental in developing the ERICA embryo scoring application. They now hold senior positions at Conceivable Life Sciences: Roberto Valencia-Murillo is Vice President of Intelligent Systems, Adolfo Flores-Saiffe Farías is Vice President of Research, and Gerardo Mendizabal-Ruiz is Vice President of Product Engineering.

for the same embryos and asked to make a final judgment combining both sources of information. Each embryologist conducted this process on between 20 to 150 simulated IVF cycles to assess performance in identifying euploid embryos.

Results: Both embryologists and ERICA demonstrated a statistically significant ability to identify aneuploidy better than random selection (both $P < 0.001$). ERICA outperformed embryologists in selecting euploid embryos on the first attempt ($P = 0.002$). No significant difference was observed between groups of embryologists in picking a better or worse embryo and, given the opportunity to change their mind in the light of the ERICA output, there was no significant change for the better or worse, despite a large number changing their minds.

Conclusion: The study highlights that AI tools like ERICA can enhance the reliability of embryo selection by reducing subjectivity and bias; however, the combination of human and AI judgment does not always provide a clear advantage over using either method independently. These findings emphasize the need to better understand the influence of AI on human decision-making and the trust placed in automated processes in IVF settings.

INTRODUCTION

With over 12 million IVF babies born since the first (Steptoe and Edwards, 1978), notable advances have improved live birth rates per embryo transfer to over 65% in some clinics (Fauser et al., 2019). These advances include enhanced culture media, vitrification, improved ovarian stimulation, optimizing endometrial receptivity and the ability to detect aneuploid embryos (Niederberger et al., 2018). Conversely, other aspects such as manual manipulation of gametes/embryos and standard light (phase contrast) microscopy have remained remarkably unchanged since IVF's inception (Campbell, 2022). Visual inspection and the subjective view of the embryologist in embryo selection has also remained the primary decision-making tool. Different classification systems have been proposed over the years and, depending on country, these usually include factors such as number of cells, division rates, equality of blastomere size and levels of fragmentation (Dawson et al., 1995). Selecting the best embryo for transfer requires expertise, but various laboratory conditions can also influence embryo morphology (Cairo Consensus Group, 2020). These factors, in turn, affect the selection process, requiring embryologists to carefully consider both experience and environmental conditions.

In the critical pursuit of defining the most suitable embryo for selection, many investigators have used different morphological evaluations from zygote patterns (Tesarik et al., 1999) to a variety of scoring systems (Sakkas et al., 1998; Balaban et al., 2011). As the IVF world progressed to blastocyst embryo transfer (rather than cleavage stage), factors such as hatching status, blastulation rate, plus assessment of inner cell mass and trophectoderm morphology are also

considered. Gardner and Schoolcraft's initial classification of blastocysts (Schoolcraft et al., 1999) has been frequently revised and challenged (Gardner et al., 2000; Veeck and Zaninovic, 2003; Balaban et al., 2006; Richardson et al., 2015; Saiz, 2018). At the time of writing, however, no uniform consensus has been reached on how to grade blastocysts (Puga Toress, 2017). In the UK, embryos are usually scored according to the 'Gardner' system with high-quality embryos (AA, AB, BA or BB) considered for transfer, whereas the fate of poor-quality embryos depend on local standard operating procedures and single embryo transfer policies that determine whether they are typically transferred, cryopreserved or discarded. Low chances of survival and likelihood of aneuploidy are the most cited reasons for not transferring low scoring embryos (Ahlström, 2011), although such embryos can lead to viable offspring (Al Hashimi et al., 2024). Many clinics modify this system with personalized scores of letters and '+' to denote preference. For example, in Spain, most clinics follow the Association for the Study of Reproductive Biology (ASEBIR a Spanish society) scoring system (Cuevas Saiz et al., 2018), which is dynamic in nature, takes into account scientific publications and classifies embryos into four categories. Now that preimplantation embryos can successfully be video recorded (Kovacs, 2014), this scoring system is again under review for future modification, incorporating key morphokinetic parameters, morphological events and development milestones (Garcia-Belda, 2024).

Timelapse imaging (TLI) and monitoring offers enriched, continuous insights and enables dynamic embryo scoring, necessitating an evolution in traditional assessment methods. With digitalization, this approach transforms the humble microscope into an advanced analytical

tool, elevating its role from mere observation to active data analysis. Timelapse imaging can output either static or video images and gives the embryologists an opportunity to observe more aspects of development. Such visualization does not require removal of the embryo from the incubator, and this uninterrupted culture plus more data-rich observation of human embryo development has the potential to improve embryo selection (Márquez-Hinojosa et al., 2022). The benefits of TLI in improving IVF outcomes are, however, still uncertain (Armstrong et al., 2019); some investigations have provided evidence that time-lapse monitoring, alongside morphokinetic algorithms for embryo selection, can result in improved clinical outcomes. The meta-analysis by Pribenszky et al. (2017) established that TLI was associated with improved ongoing clinical pregnancy rates, early pregnancy loss and live birth rates. Inconsistencies in the studies included in the meta-analysis and the selective application of the technology, however, cast doubt on the quality of the overall conclusions, as did potential conflicts of interest plus inclusion/exclusion criteria (Alikani, 2018; Armstrong et al., 2018). Fishel et al. (2018; 2020) subsequently established that morphokinetic-based algorithms could provide objective hierarchical embryo quality ranking with better discriminating power than visual inspection by an embryologist. Time-lapse systems, however, can be costly, and analysis can be time-consuming (Conaghan, 2013; Kirkegaard, 2015; Kaser, 2016; Chen, 2017; Sciorio, 2021).

The pros and cons of TLI must, therefore, be assessed on a clinic-by-clinic basis, as the debate surrounding its use shows no signs of abating. Recent controversy, particularly highlighted by Kieslinger et al. (2024) suggesting 'an inconvenient reality', implies that technologies such as TLI,

whether enhanced with artificial intelligence (AI) or not, are often subject to significant hype in their applications, citing the recent randomized controlled trial (RCT) by *Illingworth et al., (2024)*. The RCT demonstrated the non-inferiority of AI compared with manual methods of embryo ranking. The study, however, also uncovered notable real-world truths: first, the speed of AI was markedly faster than that of its embryologist counterpart, and second, manual assessments can be fluctuating. This highlights 'the Hawthorne effect', in which embryologists performed better than usual b of the study. Furthermore, technologies such as TLI, potentially supported by AI, may offer practical benefits, such as uninterrupted and continuous embryo culture, which could help alleviate the growing challenges faced by embryologists in modern clinics (*Choucair et al., 2021*).

Automated annotation of developmental milestones and the use of AI are obvious steps through which grading of embryos by visual inspection might be developed and improved (*VerMilyea, 2020a; Malmsten, 2021*). The inclusion of AI in embryo selection would be highly desirable, as its ability to make, or contribute to, decisions is reliable and reproducible, unlike humans, who can be prone to subjectivity and other bias. To date, however, its use has been somewhat limited to early developmental stages based on pre-existing classifications or making use of sophisticated microscopy modalities (Mio, 2006; *Storr, 2015; Goodman, 2016*). Some important demonstrations of the potential power of this approach have been made. For example, *Dimitriadis et al. (2019)* described the development of a convoluted neural network that could distinguish, with over 90% accuracy, two pronuclei and non-two pronuclei zygotes at 18 h after insemination, whereas *Zhao et al. (2021)* showed that assessment of the morphokinetic patterns of the zygote cytoplasm, zona pellucida and pronuclei could achieve human embryo segmentation precisely, quickly and reproducibly. Using single-image analysis, *Khosravi et al. (2019)* used an AI deep learning model to assess blastocyst quality prioritizing expansion. This was followed by inner cell mass, then trophectoderm quality, developing a blastocyst ranking system called the 'BL score', demonstrating over 50% positive outcomes. Moreover, *Kragh et al. (2019)* found a reproducible way to assess embryo quality that was at least on a par with the

judgement of the embryologist, with the AI system having slightly better correlation between predicted embryo quality and implantability than human embryologists.

In general terms, an under-studied area in blastocyst selection centres around the reasons for agreement and disagreement. This relates to similarities and differences between embryologists, especially when comparing the choices made by AI models to that of human beings. Specifically, how different is blastocyst selection for embryo transfer between embryologists using a blastocyst grading system compared with using an AI model selection? Moreover, how often will the AI choose a different blastocyst compared with the embryologist, particularly when embryologists cannot necessarily always agree, even among themselves? Currently, no standards exist for choosing an AI system for embryo evaluation. They depend on the type and size of the data set, as well as the output queries (*Fernandez, 2020*) and the specific time point for image analysis (*VerMilyea 2020b; Zaninovic 2020*).

The relationship between embryo morphology and chromosome abnormality, principally aneuploidy, has been the subject of studies, involving standard observation and TLI. In general terms, poor morphology can be indicative of chromosomal abnormality. For instance, gross chromosome abnormalities are often observed alongside differences in morphokinetics, such as embryo developmental delay and fragmentation (*Pabon et al., 1989; Almeida et al., 1995; Munne and Alikani, 2011; Sun et al., 2012*). Many factors during in-vitro culture or egg collection and handling, e.g. increased temperature and high oxygen tension, can affect cytoskeleton organization adversely, and cause spindle damage and subsequent chromosomal errors. Specifically, temperature changes in egg collection and handling have been demonstrated to cause meiotic aneuploidy (*Pickering et al., 1990; Mortimer and Mortimer, 2015*), whereas unsuitable temperature control during in-vitro culture can cause embryonic mosaicism (*Munne et al., 1997*). Nonetheless, the ability to predict aneuploidy through morphology alone is problematic (*Campbell et al., 2013; Ottolini et al., 2014*). There is, therefore, an ongoing question of whether aneuploid embryos have notably different morphokinetic properties to that of their euploid counterparts. In a systematic

review, *Bamford et al. (2022)* collated the results of 58 studies from over 40,000 embryos. They identified 10 morphokinetic features that had been implicated as being significantly altered in aneuploid compared with chromosomally normal embryos. These features included the time to reach eight cells, the time to blastulation and the time to achieve an expanded blastocyst. They also reported prognostic potential for the degree of fragmentation, for multinucleation continuing up until the four-cell stage and for the frequency of embryo contractions, but not for early multinucleation nor unequal cleavage. Notably, algorithms involving AI for live births could also have some predictive value (*Bamford et al., 2023; 2024*).

Non-selection trials have established that, when an embryo is diagnosed as fully aneuploid, i.e. 100% of the five to 10 cells in a trophectoderm biopsy all have a chromosome abnormality, then the chances of it developing to a chromosomally normal live birth are slim at best. Of the four studies in which known aneuploid embryos (three non-selection trials and one unblinded cohort study), 267 embryos were transferred with only three leading to live births, a little over 1% chance (*Tiegs et al., 2021; Wang et al., 2021; Yang et al., 2021; Barad et al., 2022*). By contrast, where all cells are detected as euploid, chances of normal live births are around 65% according to *Tiegs et al. (2021)*. The situation is clouded somewhat by euploid/aneuploid mosaicism arising post-zygotically. Mosaicism is discussed extensively in different forums elsewhere but, in general terms, chances of live birth may be reduced in mosaic embryos, depending on the proportion of aneuploid cells. The debate surrounding the relative benefits and pitfalls of preimplantation genetic testing for aneuploidy (PGT-A) is long-standing and continuing; however, the fact remains that the procedure involves an added financial burden to the patient and, if a low-cost option could be found for selecting euploid embryos, and without the need to carry out embryo biopsy, it would most likely achieve considerable take-up.

With the above in mind, the question arises whether embryologists can regularly and reliably differentiate euploid from aneuploid embryos through morphology alone, whether AI-assisted selection can perform the selection more effectively, or both. Although PGT-A remains the most objective way to assess an embryo for

aneuploidy, its invasive nature, cost and concerns about its diagnostic accuracy, e.g. with mosaicism, limit more widespread use. Although non-invasive approaches for embryo selection, e.g. time-lapse morphokinetic evaluation ([Campbell 2013](#)), morphology assessment ([Capalbo, 2014](#); [Zhan, 2020](#)) and AI systems ([Chavez-Badiola et al., 2020](#)) have aimed to compare PGT-A outcomes against their findings, it is still difficult to find studies presenting AI systems for embryo ranking trained against ploidy status as a ground truth. [Salih et al. \(2023\)](#) presented a systematic review of 20 studies and found that AI models consistently outperformed embryologists in predicting embryo quality and clinical outcomes. This highlights the potential of AI to enhance the reliability of embryo selection in IVF. The investigators, however, conclude that, although the results are promising, further research and validation are necessary before implementing AI for embryo selection in clinical practice.

Our own work ([Chavez-Badiola et al., 2020](#)) developed the AI clinical assistant 'ERICA' (Embryo Ranking Intelligent Classification Algorithm) to rank embryos based on the ability to predict euploidy and pregnancy test results. ERICA uses a single static blastocyst image in a known outcome data set. After training and validation, it was established that ERICA was more successful than both random selection and experienced embryologists in correctly identifying and ranking embryos with euploidy and the highest implantation potential. More recently our group ([Chavez-Badiola et al., 2024](#)) has established a relationship between the use of ERICA and predicting the chances of spontaneous miscarriage. That is, results supported a correlation between the risk of spontaneous miscarriage and embryo rank as determined by ERICA; the classification accuracy was 67.4%.

In the present study, the use of ERICA for embryo selection is extended, using a training interface to test a number of hypotheses: first that embryologists can differentiate euploid from aneuploid embryos significantly better than random; second that ERICA is more effective than, or at least as good as, experienced embryologists for predicting euploidy in human embryos; finally, that experienced human judgement, augmented by AI (ERICA), is the superior approach compared with either alone.

MATERIALS AND METHODS

Ethical considerations

This study was reviewed and approved by the Institutional Review Board (IRB) of New Hope Fertility Centre Mexico, which is officially registered and recognized by the Comisión Nacional de Bioética (Registration Code 09-CEI-00120170131, 14 June 2021). The IRB classified the research as low risk and non-interventional, issuing a waiver for additional ethical approval in accordance with its guidelines. This approval was recorded under reference number RA-2021-04.

Study design

The basis of this prospective study was to analyse a selection tool in which 19 professional embryologists were presented with a set of images of blastocyst embryos and asked to rank, on the basis of morphology alone, which embryos they thought were most, and least, likely to be euploid. The ERICA algorithm, a tool trained on a series of 840 prior images (independent to the images used in this exercise) from which, via deep neural networks and artificial vision, 94 features from each image were extracted and used to assess ploidy and implantation ([Chavez-Badiola et al., 2020](#)), was also assessed using the same selection tool. In addition to using the tool, the professional embryologists were also presented with the results of a previous assessment of the embryo images from ERICA to investigate if, and how, this additional information resulted in them changing their decision.

All images in the selection tool were from prior PGT-A cases carried out in 10 participant clinics (New Hope, New York City, USA; Hanabusa, San Diego, USA; New Hope, Guadalajara, Mexico; New Hope, Mexico City, Mexico; IECH, Monterrey, Mexico; Institute of Life-IASO, Athens, Greece, Poma Fertility, Kirkland, USA; AndroFert, Campinas Brazil; Aria, London, UK and IVF London, UK) between January 2020 and October 2021.

The stimulation and embryology protocols across the participating groups, though varying in the type of incubator used, adhered to standardized procedures for ovarian stimulation and oocyte retrieval ([Pacchiarotti et al., 2016](#)). The oocyte collection, fertilization and culture processes followed industry-standard laboratory protocols. Metaphase II oocytes were inseminated via intracytoplasmic

sperm injection (ICSI) ([Palermo et al., 1992](#)) and subsequently cultured in a continuous sequential culture medium for 5 to 6 days under low oxygen conditions (5% O₂) using the optimum requirements for embryo culture in IVF systems ([Cairo Consensus Group, 2020](#)).

To create the 'cases' presented in the selection tool, a database was created of blastocyst images (all day 5), their known ploidy status (from PGT-A) and metadata (patient's age, embryo age at the time of the picture and ERICA score). Exclusion criteria included poor image quality (using ERICA's standards), mosaic and unknown ploidy diagnoses. A probabilistic distribution of blastocysts was then created with known ploidy in a cycle by age groups at the time of egg retrieval, according to the following maternal age groups: younger than 35 years (euploidy rate of 59%), 35–37 years (53%), 38–40 years (44%), 41–42 (31%), over 42 years (19%), then ran the following process ([FIGURE 1](#)) to create the synthetic embryo scaffolds: collection of one cycle size from the corresponding distribution that is between 2 and 6; collation of the number of euploids within the cycle corresponding to variations in chromosome abnormalities in human blastocysts according to age ([Sawarkar et al., 2021](#)); if the cycle contained at least one euploid and one aneuploid, the process was continued or else the first step restarted; according to their ploidy status, we assigned randomly without replacement with real embryo cases (images and metadata) from our database of embryos with known ploidy. This process was iteratively repeated until 30 synthetic cycle scaffolds had been created from each age group, for a total of 150 cases.

In each exercise, five cases of two to five blastocyst images were presented, and the embryologists were limited to only these five cases per day to minimize operator fatigue. They were given the opportunity to assess 150 cases in total, i.e. 30 individual days at 5 cases a day. In short, embryologists had to select the best embryo from the available cohort, even when all embryos were of only fair to medium quality.

In the interface, users were asked to choose the embryo that they believed was most likely to be euploid ('best'), then most likely to be aneuploid ('poorest'). If more images remained, then the next most likely to be euploid, the next mostly likely to be

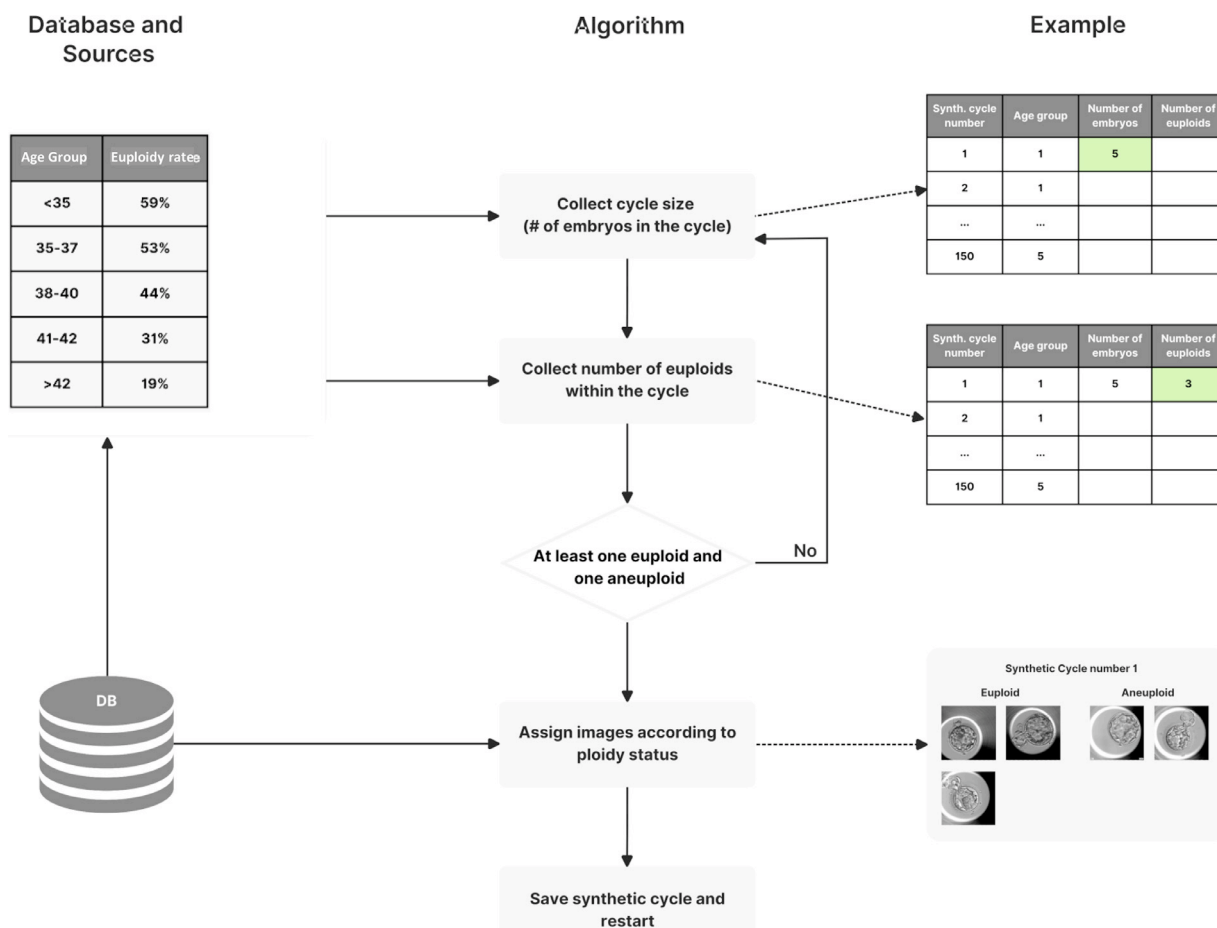


FIGURE 1 Synthetic cycle creation process by which the ‘cases’ presented in the selection tool were created. DB = database.

aneuploid were chosen until no images were left (FIGURE 2A, with an example of a full exercise shown in Supplementary Figure 1). This was achieved through a series of ‘drag and drop’ exercises, the user alternatively being encouraged to select the ‘best’ and ‘poorest’ embryo. Next, the user was encouraged to select ‘rank with ERICA’ and then informed whether AI (ERICA) had previously assigned the embryos as ‘optimal’ ‘good’ ‘fair’ or ‘poor’. The penultimate stage then encouraged the user to combine their prior judgment with ERICA’s assessment and select an embryo for transfer (FIGURE 2B). In the final stage, the user was informed whether the embryo was euploid or aneuploid. If the former (euploid), the exercise ended for that cycle (with the message ‘well done, the embryo is euploid’), if the latter (aneuploid), then they were encouraged to select another embryo (with the message ‘the embryo is aneuploid, try again’) until a euploid one had been selected. Each cycle had at least one (but no more than two) euploid embryo(s) and one to four aneuploid

embryos, with the number of euploid embryos in the case presented in the selection tool.

Outcome measures were as follows: mean number of attempts to select a euploid embryo by the user alone without ERICA’s assistance among all cycles attempted with 1 being the best possible score; the proportion of euploid embryos selected at the first try alone without ERICA’s assistance; the rate of opinion change, measured as the proportion of cycles where, following ERICA’s assessment, the user decided to transfer an embryo that they did not originally select as their first choice; the rate of opinion change for the better, measured as the proportion of opinion changes where the original selection was aneuploid but, after ERICA’s assessment, the user selected a euploid embryo; and the rate of opinion change for worse, measured as the proportion of opinion changes in which the original user selection was euploid but, after ERICA’s assessment, the user selected an aneuploid embryo.

Finally, all users were triaged into three groups as follows: ‘junior embryologist’ (less than 1 year’s experience and less than 100 embryos assessed in their career [$n = 6$]); ‘intermediate level embryologist’ (1–5 years’ experience and 100–600 embryos assessed); and ‘experienced embryologist’ (more than 5 years’ experience and more than 600 embryos assessed [$n = 7$]). Junior embryologists were all proficient at embryo assessments and transfers; intermediate embryologists were, in addition, proficient at cryopreservation and ICSI; experienced embryologists were also proficient at performing embryo biopsy ($n = 6$). Simulated, random, data were also created by running the selection tool with 30 ‘virtual, random’ users performing the exercise, an average of 90 times with no selective judgement. The six junior embryologists all carried out 150 simulated cycles (total 900, mean 150); the six intermediate embryologists carried out 47, 58, 140, 150, 150 and 150 (total 695, mean 115.8) simulated cycles, respectively; the seven experienced embryologists carried

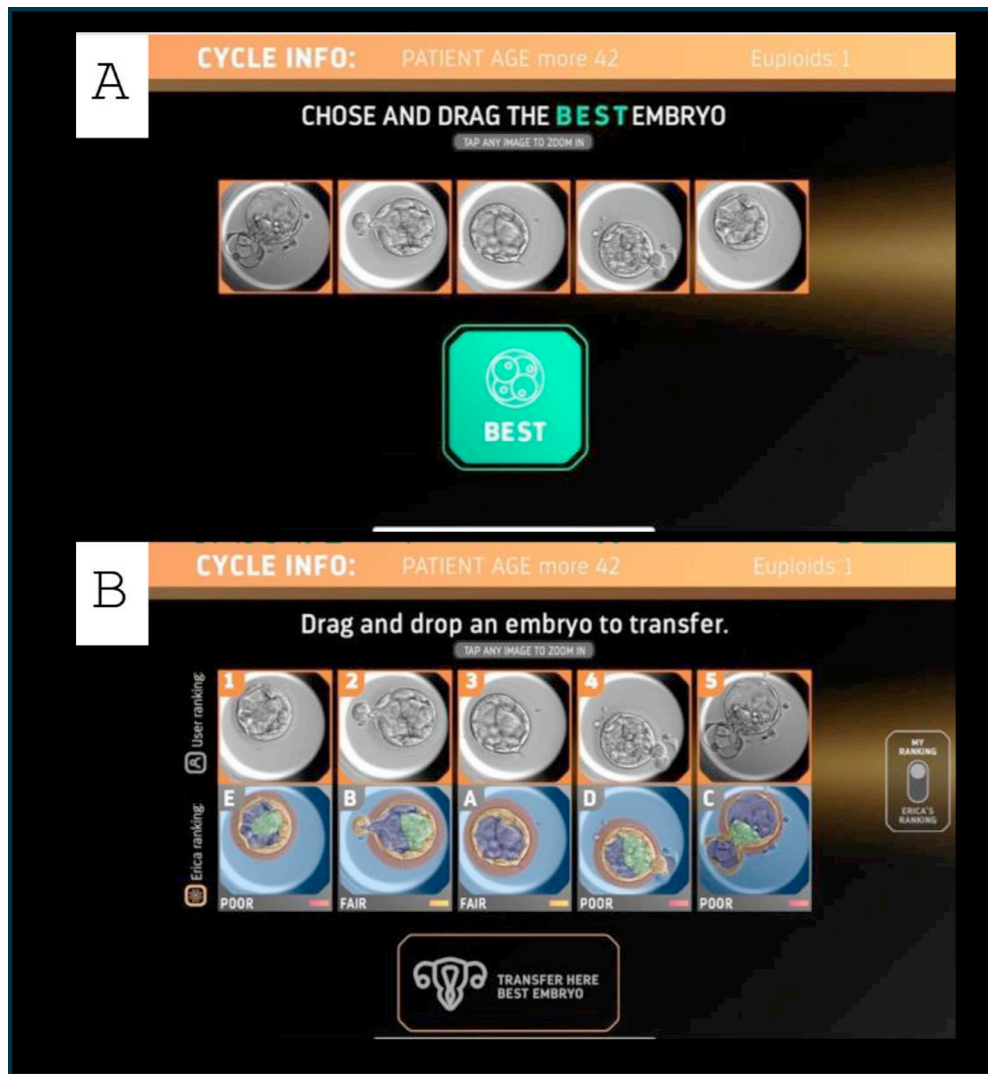


FIGURE 2 Examples of the selection tool. In the first selection step of each exercise (A), users are asked to assess the embryo images presented and to 'drag and drop' the 'best' embryo onto the selection box. In the final selection step (B), users are asked to select an embryo to transfer.

out 20, 20, 149, 150, 150, 150 and 150 (total 789, mean 112.7) simulated cycles, respectively.

Data were analysed using Kruskal–Wallis tests with, where appropriate, Bonferroni corrected Dunn tests used for post hoc testing, in R version 4.2.2 (R Core Team, 2022) using RStudio (RStudio Team, 2020).

RESULTS

Comparison of the proportion of euploid embryos selected at the first try indicates that the groups differed (FIGURE 3A) (Kruskal–Wallis chi-squared = 64.612, $df = 2$, $P = 9.327 \times 10^{-15}$). Here, ERICA performed marginally better than the embryologists (post hoc $P = 0.002$), and both ERICA and the embryologists

performed better than randomly (post hoc $P < 0.001$), selecting euploid embryos on the first try 52 and 49% of the time, respectively, compared with 40% of the time for the randomizer (FIGURE 3A). Similarly, analysis of the number of attempts to select a euploid embryo also indicated that the groups performed differently (FIGURE 3B) (Kruskal–Wallis chi-squared = 36.091, $df = 2$, $P = 1.455 \times 10^{-8}$), with the embryologists and ERICA again outperforming the randomizer (post hoc $P < 0.001$). For this measure, the embryologists and ERICA needed an average of 1.77 and 1.73 attempts to reach a euploid conceptus, respectively, compared with 1.91 for the randomizer. Comparing experienced embryologists' performance against those less experienced (FIGURE 4) any observed differences were not statistically significant

for the proportion of euploid embryos selected at the first try or the number of attempts to select a euploid embryo (Kruskal–Wallis chi-squared = 0.16748, $df = 2$, $P = 0.92$ and Kruskal–Wallis chi-squared = 1.4176, $df = 2$, $P = 0.49$, respectively).

When presented with the assessment of the embryos from ERICA, a large proportion of all career stages of embryologists changed their decisions. Overall, the rate of decision change did not differ between groups (FIGURE 5A): Kruskal–Wallis chi-squared = 0.56986, $df = 2$, $P = 0.7521$. No significant difference was found between groups of embryologists in picking a better or worse embryo (FIGURE 5B): Kruskal–Wallis chi-squared = 1.0682, $df = 2$, $P = 0.5862$ for the rate of decision change for the better,

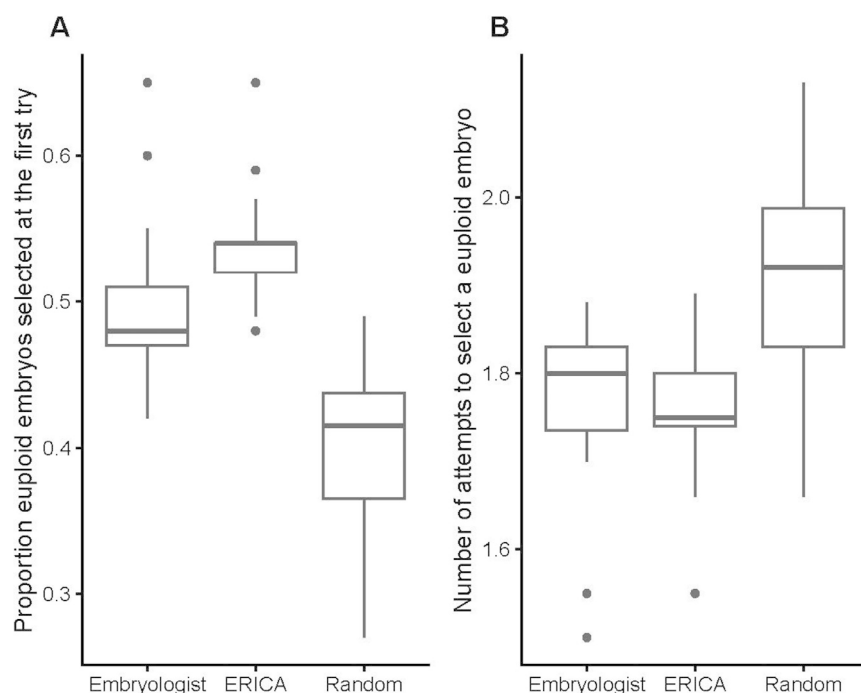


FIGURE 3 Embryologists and ERICA perform better than random. Box and whisker plots showing (A) the proportion of euploid embryos selected at the first try for embryologists based on morphology alone, ERICA and the randomizer, with post-hoc testing by Bonferroni corrected Dunn test indicating that ERICA outperforms both the embryologists and the randomizer ($P = 0.002$ and $P = 1.629793e-15$, respectively), and that the embryologists outperform the randomizer ($P = 9.400497e-04$); (B) the number of attempts to select a euploid embryo for embryologists, ERICA and the randomizer, with post-hoc testing by Bonferroni corrected Dunn test indicating that both the embryologists and ERICA outperform the randomizer ($P = 2.952989e-04$ and $P = 4.413281e-09$, respectively), whereas no difference was found between the embryologists and ERICA ($P = 0.44$). Box and whisker plots show median (interquartile range), maximum and minimum and outliers. ERICA and the randomizer performed 150 cycles each. The six junior embryologists all carried out 150 simulated cycles; the six intermediate embryologists carried out 47, 58, 140, 150, 150 and 150 simulated cycles, respectively; the seven experienced embryologists carried out 20, 20, 149, 150, 150, 150 and 150 simulated cycles, respectively.

and **FIGURE 5C**: Kruskal–Wallis chi-squared = 0.80044, $df = 2$, $P = 0.6702$ for the rate of decision change for the worse.

DISCUSSION

In the present study, we accept our hypothesis that both human operators and ERICA can differentiate euploid from aneuploid embryos significantly better than random. We provide evidence that ERICA is at least as good as embryologists using morphology alone in number of attempts to select a euploid embryo. We provide evidence that ERICA performs better than embryologists using morphology alone in choosing a euploid embryo on the first attempt. Finally, we reject the hypothesis that experienced human judgement, augmented by AI (ERICA), is the superior approach compared with either approach alone as there was no significant ‘change for the

better’ nor significant ‘change for the worse’ when embryologists were given the opportunity to change their mind in the light of the ERICA decision. The decision to change data are, however, suggestive that senior embryologists may be less likely to change their decisions and any changes they do make may be less likely to be for the worse, although the differences are not significant. Larger numbers would be required to establish whether this is the case.

Clinical embryology is relatively unique as a medical laboratory profession in that it has limited automation; it relies heavily on manual dexterity, teamwork and subjective judgment to produce optimum results (*Campbell et al., 2022*). Moreover, the role of the embryologist has become more complex in recent years, with additional duties and responsibilities, including the increase of administrative duties, increased cryopreservation and embryo biopsies

being performed more routinely (*Choucair 2021; Rienzi et al., 2021*).

One potentially confounding factor that can affect AI protocols is that the morphokinetic annotations themselves are initially assigned by humans and are thus subjective. Developing AI models to recognize abnormal karyokinetic (nuclear) and cytokinetic abnormalities, e.g. direct divisions one to three cell fusion, will be necessary for optimal automatic annotation. Most studies have used retrospective data under experimental settings, and the clinical application of AI still requires prospective investigations. In studies using AI to predict embryo implantation potential on static or time lapse images, secondary factors, such as laboratory conditions or other human factors, have not been analysed, nor included in the models. Culture conditions and human expertise are important factors that influence embryo development and quality. These factors will need to be incorporated into the models to achieve a useful and objective prediction. In addition, we know that successful implantation and live birth depend on other factors not inherent to the embryo. Predicting implantation solely on embryo quality is an incomplete assessment. AI embryo prediction models should focus on ranking the embryos within the patient cohort rather than on implantation prediction. The variation in success rates among IVF centres and laboratories prevents the establishment of universal AI models for implantation prediction. A further confounding factor is that, although, the number in each group (6, 6 and 7) was roughly equal, the number of cases performed (900, 695 and 789), and mean cases performed (150, 115.8, 112.7) was greater in the junior embryologists group. This might imply that a slightly greater ‘weighting’ should be given to the results of the junior group. This is not something we included here; however, it will be a consideration for future studies.

Although RCTs remain the gold standard in medical research (*Grossman et al., 2005*), their ability to fully assess the effect of AI in embryology is unclear. The lack of trials may slow adoption, as clinics weigh cost–benefit concerns. The significant reduction in annotation time with the use of AI, however, highlighted in the recent RCT by *Illingworth et al. (2024)*, suggests its value may lie in efficiency rather than clinical superiority. Although larger trials are needed, AI’s speed advantage could

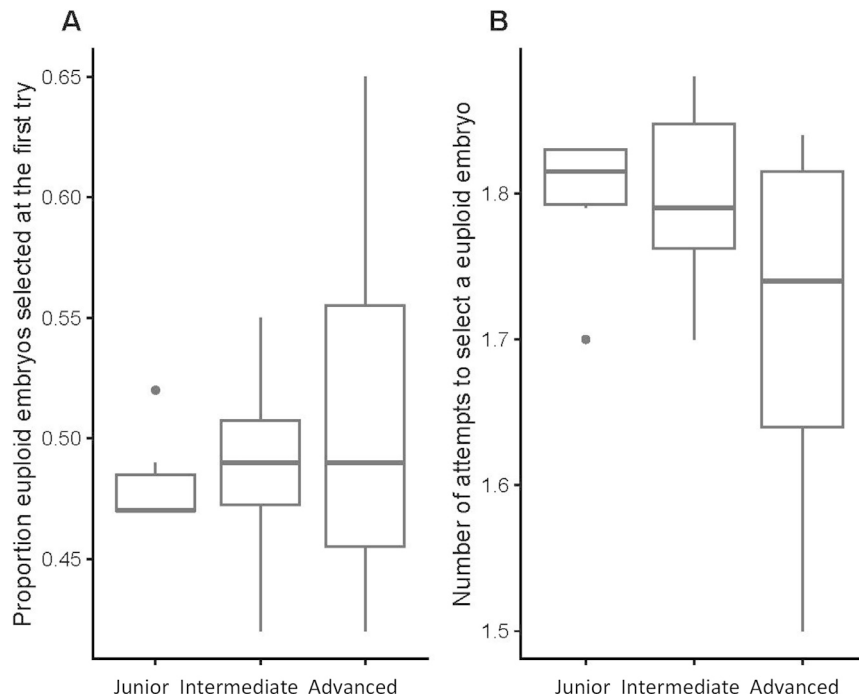


FIGURE 4 Box and whisker plots showing (A) proportion of euploid embryos selected at the first try and (B) the number of attempts to select a euploid embryo (Kruskal–Wallis chi-squared = 0.16748, $df = 2$, $P = 0.92$ and Kruskal–Wallis chi-squared = 1.4176, $df = 2$, $P = 0.49$, respectively), based on morphology alone. Box and whisker plots show median (interquartile range), maximum and minimum and outliers. The six junior embryologists all carried out 150 simulated cycles; the six intermediate embryologists carried out 47, 58, 140, 150, 150 and 150 simulated cycles respectively; the seven experienced embryologists carried out 20, 20, 149, 150, 150, 150 and 150 simulated cycles, respectively. Although senior embryologists showed a trend towards better performance than those less experienced (junior and intermediate), these differences were not statistically significant.

drive its integration into fertility laboratories.

The question still remains, however: how might AI predictions influence the decision-making process of clinical embryologists? The research herein presented sets out to quantify this with a simulation that uses embryos of known ploidy in a ‘game-like’ scenario. This ‘gamification’, while important to deduce how AI influences choice for embryo selection, also acts as a training exercise, familiarizing the operator with the software that can be used in clinical settings and educating themselves to the dilemma of choice.

We anticipate that other similar studies will follow shortly, presenting new approaches aimed at embryo selection based on ploidy. These studies will perhaps target time-lapse sequences (Barnes, 2020) and incorporate omics (Bori, 2021), non-invasive chromosome screening tests (Chavez-Badiola, 2020), as well as new AI approaches. Building high-quality datasets

from diverse settings, while managing hype (VerMilyea, 2020b) and expectations are challenges that are ahead of us. Evidence seems to be accumulating in support of using AI for embryo selection. Loewke et al. (2022) conducted a retrospective study evaluating an AI model for ranking blastocyst-stage embryos. The study demonstrated using the performance metric area under the curve that the AI model has the potential to improve clinical pregnancy predictions compared with manual morphology grading, although it also identified limitations related to image quality, bias and granularity of scores that may affect clinical use.

Recently, an international survey (Palmer et al., 2024) demonstrated that most embryologists (93.7%) are well-informed about AI in the workplace. Moreover, despite recognizing its limitations, they are open to its adoption. Tools that support decision-making could provide valuable assistance in daily tasks; most embryologists view the use of AI in the laboratory as inevitable. They express a

willingness to incorporate AI in the selection of gametes and embryos to enhance clinical outcomes. According to the survey, 59.8% respondents believe AI should be integrated into routine laboratory procedures, and 73.3% are open to adopting AI technologies in the near future. This enthusiasm, however, is tempered by a desire for more evidence of AI’s effectiveness, with only 5.4% of respondents feeling extremely confident in using AI in their roles, indicating a need for more training and experience. The gamification of such applications may be a way to increase uptake through familiarity and trust of such AI tools.

To our knowledge, this is the first paper to use an app incorporating ‘simulated cycles’ to compare embryologists’ decisions with those of an AI algorithm. Trust will be a key factor if AI is to gain widespread adoption in IVF laboratories. Embracing technological advancements is essential as fertility clinics face increasing complexity and workload demands. This approach aligns with efforts to ‘democratize IVF’, a vision promoted by advocates such as Sable et al. (2023), who argue that technology can significantly enhance accessibility and efficiency in the fertility sector (Saleh et al., 2023). Trust remains an issue to health professionals using AI (Tucci, 2022). The lack of trust in the AI output balanced with the embryologists’ experience introduces challenges and barriers to adoption and implementation into clinical practice. To conduct extensive training on AI with tools such as ERICA X (the iOS/android mobile phone application that is a downloadable version of the clinic specific ERICA software) will help to understand the trust dynamics between AI and expert embryologist-end-users. Trust enables us to manage uncertainty, especially in interactions with AI systems. Trust in AI, however, is shaped by human factors, such as user education, biases and perceptions, along with the system’s characteristics, such as ease of use and transparency. Among these, reliability is critically defined as the AI’s ability to perform tasks predictably and consistently, which significantly impacts user trust (Asan, 2020).

Ethical considerations using AI also remain central to the field of embryology, with 45% of surveyed embryologists expressing at least some concern about these issues (Palmer et al., 2024). The WHO (2024), in a recent technical draft, echoes these concerns,

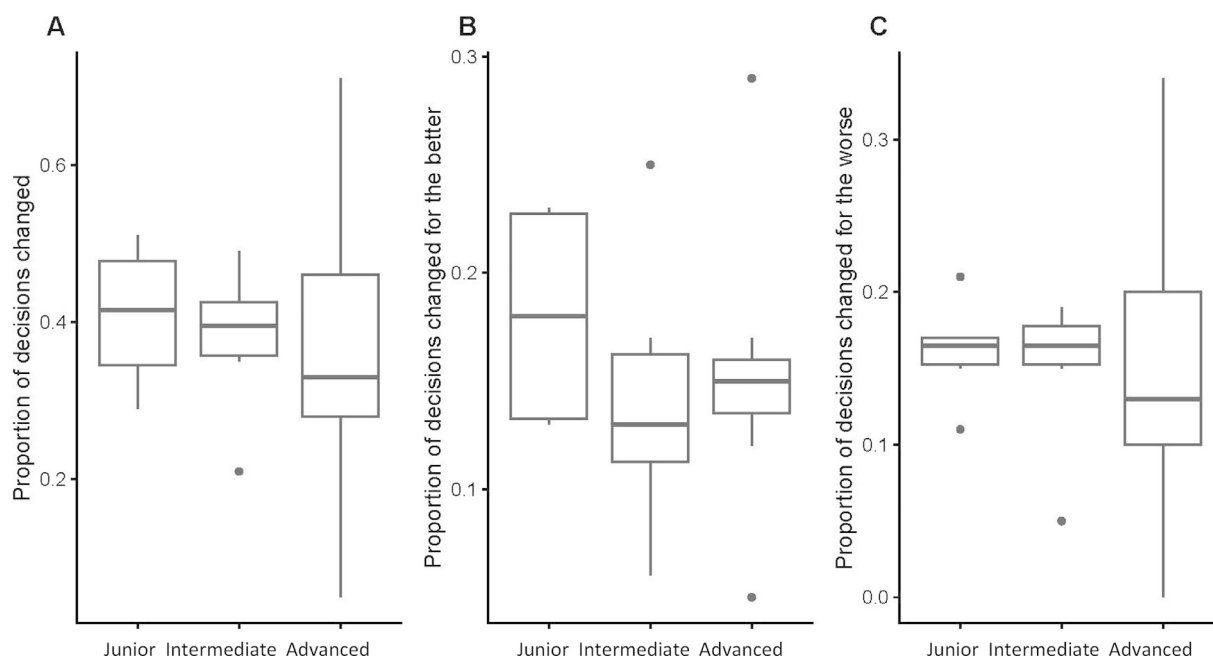


FIGURE 5 Box and whisker plots showing decisions when presented with the assessment of the embryos from ERICA; a large proportion of all career stages of embryologists changed their decisions. (A) Kruskal–Wallis chi-squared = 0.56986, $df = 2$, $P = 0.7521$). The six junior embryologists all carried out 150 simulated cycles; the six intermediate embryologists carried out 47, 58, 140, 150, 150 and 150 simulated cycles, respectively; the seven experienced embryologists carried out 20, 20, 149, 150, 150, 150 and 150 simulated cycles, respectively. No difference was found between groups of embryologists in picking a better or worse embryo; (B) Kruskal–Wallis chi-squared = 1.0682, $df = 2$, $P = 0.5862$ for the rate of decision change for the better; and (C) Kruskal–Wallis chi-squared = 0.80044, $df = 2$, $P = 0.6702$ for the rate of decision change for the worse). Box and whisker plots show median (interquartile range), maximum and minimum and outliers.

addressing the potential risks and benefits associated with advancing technology in healthcare. This document underscores the need for caution to mitigate biases and minimize potential harms, particularly with the integration of AI and automated tools in clinical settings. As data-driven, automated systems become more integrated into decision-making, it is critical for users to understand when to rely on these tools and when human judgment should take precedence (Mozannar et al., 2021). In addition, in many places in the world, particularly the USA, concerns over potential litigation from patients may lead some clinics to store or use embryos purely and primarily as a result of competitive practices among fertility centres, as rival clinics may offer comparable embryo services, potentially opening the door to lawsuits (Kitts, 2023; Gulrajani, 2023). Standardized approaches, aided by AI, may help democratize the decision-making process of when to store and when to use embryos. A further factor is operator fatigue. In another survey of US embryologists focusing on this issue (Murphy, 2022) it was shown that automation was quoted as one of the

factors that embryologists feel will reduce fatigue and eliminate duplication.

AI systems in medicine tend to be complex and unpredictable. They can lack evidence, be difficult to grasp, and the many uncertainties and risks related to their use, such as potential patient harm are often cited (ERPS, 2022). Although the media often portrays AI in medicine with optimism, public opinion tends to be more cautious and even distrustful. This gap underscores the need for transdisciplinary training programmes that integrate computer science with medical education (Yakar, 2022). Trustworthiness in AI systems is essential for their successful adoption in healthcare, and one effective approach to building this trust is through hands-on familiarity and training, as exemplified in this study. Here, through an interactive training tool, professionals learn to work harmoniously with AI-driven software. We present a means to introduce the concepts of AI in the IVF laboratory and to provide familiarity that would lead to better adoption of the tool in real life. Indeed, the interplay between AI and human decision-making merits further examination, ideally using a more extensive dataset to yield more comprehensive

insights. Simply considering outside influence there may be ‘double guessing judgements’ or ‘overthinking the embryological decision’ leading to reduced clinical effectiveness and less potential. This has been seen in medicine when clinicians are influenced by peers (Bulls, 2024) or by being under review (Anderson, 2021). Such metacognition, the ability to reflect on their thought process and choose an effective strategy, and self-efficacy have served embryologists well up to now. They are nonetheless being challenged by the advent of AI ranking tools.

In clinical practice, embryologists consider additional contextual factors, such as embryo freezing dates and maternal age, when selecting embryos. These elements provide valuable insights that could enhance decision-making, especially across multiple IVF cycles. Future models incorporating such context may improve predictions and better reflect clinical complexity. This study focused on ranking embryos based on implantation potential, independent of patient-specific variables. By isolating morphological assessment and PGT-A results, it ensured rankings reflected

intrinsic embryo quality. Although AI models such as ERICA may have a positive bias in this task, integrating contextual variables could further enhance predictive accuracy and clinical relevance.

The present study assumes a level playing field between embryologists and AI; however, we recognize that embryologists are accustomed to specific imaging formats and contrast lighting techniques unique to their clinic, which may influence their ability to assess unfamiliar images. Despite efforts to standardize images, differences in formats remain a potential confounding factor. Future research could explore clinic-specific imaging adaptations to refine comparisons. Additionally, we acknowledge that real-world embryo selection often involves limited choices, reduces decision-making complexity and potentially favours AI. Future model iterations could integrate simpler scenarios to better reflect clinical reality. ERICA's ability to generalize across datasets offers a clear advantage; however, embryologists rely on experience and pattern recognition. Expanding standardized training sets could help embryologists refine assessments beyond their own laboratory conditions, ensuring continued advancements in embryo selection.

In conclusion, AI simulation training could play a valuable role in embryology, similar to its use in other fields like surgery ([Park et al., 2022](#)). By recreating real-life scenarios that can be endlessly repeated and refined, these medical simulations enhance skill development without replacing human expertise. Instead, they provide a controlled environment for operators to practice and improve their techniques. Humans, like AI, will improve performance with greater exposure to more data.

In conclusion, embryologists must learn to interact effectively with AI systems to optimize outcomes in the same way as has been reported in other medical disciplines. A recent study by [Goh et al. \(2024\)](#) investigating decisions made by clinicians using AI enhanced large language models demonstrated that large language models alone outperformed clinicians, even when working collaboratively. The study shows that improved interaction and understanding between human operators and AI systems are essential for achieving optimal performance, emphasizing the

need for AI to function as a co-pilot alongside human expertise. In the present study, we present a means by which this may be achieved, integrating AI into the world of embryo selection. Combining ERICA with human intervention is a tool that has the potential to be both effective and popular, engendering trust and addressing issues such as fatigue, as well as several ethical challenges. It could cut down on overthinking and could alternatively encourage strategies for self-reflection.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.rbmo.2025.104990](https://doi.org/10.1016/j.rbmo.2025.104990).

REFERENCES

- Ahlström, A., Westin, C., Reiser, E., Wikland, M., Hardarson, T., 2011. Trophoctoderm morphology: an important parameter for predicting live birth after single blastocyst transfer. *Human reproduction* 26 (12), 3289–3296.
- Al Hashimi, B., Harvey, S., Harvey, K., Linara-Demakakou, E., Raikundalia, B., Green, O., Griffin, D., Ahuja, K., Macklon, N., 2024. Reassessing the conventional fertilisation check: leveraging PGT-A to increase the number of transferrable embryos. *Reproductive BioMedicine Online*, 104595.
- Almeida, P.A., Bolton, V.N., 1995. The effect of temperature fluctuations on the cytoskeletal organisation and chromosomal constitution of the human oocyte. *Zygote* 3 (4), 357–365.
- Anderson, E.S., Griffiths, T.R., Forey, T., Wobi, F., Norman, R.I., Martin, G., 2021. Developing Healthcare Team Observations for Patient Safety (HTOPS): senior medical students capture everyday clinical moments. *Pilot and Feasibility Studies* 7, 1–8.
- Armstrong, S., Bhide, P., Jordan, V., Pacey, A., Farquhar, C., 2018. Time-lapse systems for ART. *Reproductive Biomedicine Online* 36 (3), 288–289.
- Armstrong, S., Bhide, P., Jordan, V., Pacey, A., Marjoribanks, J., Farquhar, C., 2019. Time-lapse systems for embryo incubation and assessment in assisted reproduction. *Cochrane Database of Systematic Reviews*(5).
- Asan, O., Bayrak, A.E., Choudhury, A., 2020. Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of medical Internet research* 22 (6), e15154.
- Balaban, B., Brison, D., Calderon, G., Catt, J., Conaghan, J., Cowan, L., Ebner, T., Gardner, D., Hardarson, T., Lundin, K., Magli, M.C., 2011. Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Reproductive biomedicine online*. 22 (6), 632–646.
- Balaban, B., Yakin, K., Urman, B., 2006. Randomized comparison of two different blastocyst grading systems. *Fertility and sterility* 85 (3), 559–563.
- Bamford, T., Barrie, A., Montgomery, S., Dhillon-Smith, R., Campbell, A., Easter, C., Coomarasamy, A., 2022. Morphological and morphokinetic associations with aneuploidy: a systematic review and meta-analysis. *Human reproduction update* 28 (5), 656–686.
- Bamford, T., Smith, R., Easter, C., Dhillon-Smith, R., Barrie, A., Montgomery, S., Campbell, A., Coomarasamy, A., 2023. Association between a morphokinetic ploidy prediction model risk score and miscarriage and live birth: a multicentre cohort study. *Fertility and Sterility* 120 (4), 834–843.
- Bamford, T., Smith, R., Young, S., Evans, A., Lockwood, M., Easter, C., Montgomery, S., Barrie, A., Dhillon-Smith, R., Coomarasamy, A., Campbell, A., 2024. A comparison of morphokinetic models and morphological selection for prioritizing euploid embryos: a multicentre cohort study. *Human reproduction* 39 (1), 53–61.
- Barad, D.H., Albertini, D.F., Molinari, E., Gleicher, N., 2022. IVF outcomes of embryos with abnormal PGT-A biopsy previously refused transfer: a prospective cohort study. *Human Reproduction* 37 (6), 1194–1206.
- Bulls, H.W., Hamm, M., Wasilewski, J., Olejniczak, D., Bell, S.G., Liebschutz, J.M., 2024.

- "To prescribe or not to prescribe, that is the question": Perspectives on opioid prescribing for chronic, cancer-related pain from clinicians who treat pain in survivorship. *Cancer*.
- Cairo Consensus Group, 2020. 'There is only one thing that is truly important in an IVF laboratory: everything.' Cairo Consensus Guidelines on IVF Culture Conditions. *Reproductive BioMedicine Online* 40 (1), 33–60.
- Campbell, A., Cohen, J., Ivani, K., Morbeck, D., Palmer, G., Mortimer, S., 2022. The in vitro fertilization laboratory: teamwork and teaming. *Fertility and Sterility* 117 (1), 27–32.
- Campbell, A., Fishel, S., Bowman, N., Duffy, S., Sedler, M., Thornton, S., 2013. Retrospective analysis of outcomes after IVF using an aneuploidy risk model derived from time-lapse imaging without PGS. *Reproductive biomedicine online* 27 (2), 140–146.
- Capalbo, A., Rienzi, L., Cimadomo, D., Maggiulli, R., Elliott, T., Wright, G., Nagy, Z.P., Ubaldi, F.M., 2014. Correlation between standard blastocyst morphology, euploidy and implantation: an observational study in two centers involving 956 screened blastocysts. *Human reproduction* 29 (6), 1173–1181.
- Chavez-Badiola, A., Flores-Saiffe-Farías, A., Mendizabal-Ruiz, G., Drakeley, A.J., Cohen, J., 2020. Embryo Ranking Intelligent Classification Algorithm (ERICA): artificial intelligence clinical assistant predicting embryo ploidy and implantation. *Reproductive biomedicine online* 41 (4), 585–593.
- Chavez-Badiola, A., Farías, A.F., Mendizabal-Ruiz, G., Silvestri, G., Griffin, D.K., Valencia-Murillo, R., Drakeley, A.J., Cohen, J., 2024. Use of artificial intelligence embryo selection based on static images to predict first-trimester pregnancy loss. *Reproductive BioMedicine Online* 49 (2), 103934.
- Chen, M., Wei, S., Hu, J., Yuan, J., Liu, F., 2017. Does time-lapse imaging have favorable results for embryo incubation and selection compared with conventional methods in clinical in vitro fertilization? A meta-analysis and systematic review of randomized controlled trials. *PloS one* 12 (6), e0178720.
- Choucair, F., Younis, N., Hourani, A., 2021. The value of the modern embryologist to a successful IVF system: revisiting an age-old question. *Middle East Fertility Society Journal* 26 (1), 15.
- Conaghan, J., Chen, A.A., Willman, S.P., Ivani, K., Chenette, P.E., Boostanfar, R., Baker, V.L., Adamson, G.D., Abusief, M.E., Gvakharina, M., Loewke, K.E., 2013. Improving embryo selection using a computer-automated time-lapse image analysis test plus day 3 morphology: results from a prospective multicenter trial. *Fertility and sterility* 100 (2), 412–419.
- Dawson, K.J., Conaghan, J., Ostera, G.R., Winston, R.M., Hardy, K., 1995. Delaying transfer to the third day post-insemination, to select non-arrested embryos, increases development to the fetal heart stage. *Human Reproduction* 10 (1), 177–182.
- Dimitriadis, I., Bormann, C.L., Kanakasabapathy, M.K., Thirumalaraju, P., Gupta, R., Pooniwal, R., Souter, I., Rice, S.T., Bhowmick, P., Shafiee, H., 2019. Deep convolutional neural networks (CNN) for assessment and selection of normally fertilized human embryos. *Fertility and Sterility* 112 (3), e272.
- European Parliamentary Research Service (EPRS), 2022. Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU\(2022\)729512_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf). (Accessed November 22, 2022).
- Fausser, B.C., 2019. Towards the global coverage of a unified registry of IVF outcomes. *Reproductive biomedicine online* 38 (2), 133–137.
- Fernandez, E.I., Ferreira, A.S., Cecílio, M.H., Chéles, D.S., de Souza, R.C., Nogueira, M.F., Rocha, J.C., 2020 Oct. Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data. *Journal of assisted reproduction and genetics* 37 (10), 2359–2376.
- Fishel, S., Campbell, A., Montgomery, S., Smith, R., Nice, L., Duffy, S., Jenner, L., Berrisford, K., Kellam, L., Smith, R., Foad, F., 2018. Time-lapse imaging algorithms rank human preimplantation embryos according to the probability of live birth. *Reproductive biomedicine online* 37 (3), 304–313.
- Fishel, S., Campbell, A., Foad, F., Davies, L., Best, L., Davis, N., Smith, R., Duffy, S., Wheat, S., Montgomery, S., Wachter, A., 2020. Evolution of embryo selection for IVF from subjective morphology assessment to objective time-lapse algorithms improves chance of live birth. *Reproductive biomedicine online* 40 (1), 61–70.
- García-Belda, A., Cairó, O., Martínez-Moro, Á., Cuadros, M., Pons, M.C., de Mendoza, M.V., Delgado, A., Rives, N., Carrasco, B., Cabello, Y., Figueroa, M.J., 2024. Considerations for future modification of The Association for the Study of Reproductive Biology embryo grading system incorporating time-lapse observations. *Reproductive Biomedicine Online* 48 (1).
- Gardner, D.K., Lane, M., Stevens, J., Schlenker, T., Schoolcraft, W.B., 2000. Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertility and sterility* 73 (6), 1155–1158.
- Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J.A., Kanjee, Z., Parsons, A.S., Ahuja, N., Horvitz, E., 2024. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Network Open* 7 (10), e2440969.
- Goodman, L.R., Goldberg, J., Falcone, T., Austin, C., Desai, N., 2016. Does the addition of time-lapse morphokinetics in the selection of embryos for transfer improve pregnancy rates? A randomized controlled trial. *Fertility and Sterility* 105 (2), 275–285.
- Grossman, J., Mackenzie, F.J., 2005. The randomized controlled trial: gold standard, or merely standard? *Perspectives in biology and medicine* 48 (4), 516–534.
- Illingworth, P.J., Venetis, C., Gardner, D.K., Nelson, S.M., Berntsen, J., Larman, M.G., Agresta, F., Ahtan, S., Ahlström, A., Cattrall, F., Cooke, S., 2024. Deep learning versus manual morphology-based embryo selection in IVF: a randomized, double-blind noninferiority trial. *Nature Medicine* 1–7.
- Kaser, D.J., Bormann, C.L., Missmer, S.A., Farland, L.V., Ginsburg, E.S., Racowsky, C., 2016. Eeva™ pregnancy pilot study: a randomized controlled trial of single embryo transfer (SET) on day 3 or day 5 with or without time-lapse imaging (TLI) selection. *Fertility and Sterility* 106 (3), e312.
- Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J.E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, L.A., Hickman, C., Meseguer, M., 2019. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ digital medicine* 2 (1), 21.
- Kieslinger, D.C., Lambalk, C.B., Vergouw, C.G., 2024 Oct. The inconvenient reality of AI-assisted embryo selection in IVF. *Nature Medicine*. 1, 1–2.
- Kirkegaard, K., Ahlström, A., Ingerslev, H.J., Hardarson, T., 2015. Choosing the best embryo by time lapse versus standard morphology. *Fertility and sterility* 103 (2), 323–332.
- Kitts, M., 2023. Throwing the Embryo out with the Bathwater? A Novel Evaluation of the Value of Embryos. *Journal of Applied Philosophy*.
- Kragh, M.F., Rimestad, J., Berntsen, J., Karstoft, H., 2019. Automatic grading of human blastocysts from time-lapse imaging. *Computers in biology and medicine* 115, 103494.
- Loewke, K., Cho, J.H., Brumar, C.D., Maeder-York, P., Barash, O., Malmsten, J.E., Zaninovic, N., Sakkas, D., Miller, K.A., Levy, M., VerMilyea, M.D., 2022. Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. *Fertility and sterility* 117 (3), 528–535.
- Malmsten, J., Zaninovic, N., Zhan, Q., Rosenwaks, Z., Shan, J., 2021 Apr. Automated cell division classification in early mouse and human embryos using convolutional neural networks. *Neural Computing and Applications* 33, 2217–2228.
- Márquez-Hinojosa, S., Noriega-Hoces, L., Guzmán, L., 2022 Jul. Time-Lapse Embryo culture: A better understanding of embryo development and clinical application. *JBRA assisted reproduction* 26 (3), 432.
- Mortimer, S.T., Mortimer, D., 2015. *Quality and risk management in the IVF laboratory*. Cambridge University Press.
- Munné, S., Alikani, M., 2011. Culture-induced chromosome abnormalities: the canary in the mine. *Reproductive BioMedicine Online* 22 (6), 506–508.
- Munné, S., Magli, C., Adler, A., Wright, G., De Boer, K., Mortimer, D., Tucker, M., Cohen, J., Gianaroli, L., 1997. Treatment-related chromosome abnormalities in human embryos. *Human reproduction (Oxford, England)* 12 (4), 780–784.
- Murphy, A., Baltimore, H., Lapczynski, M.S., Proctor, Jr, G., Meyer, E.C., Glynn, T., Domar, A.D., Collins, M.G., 2022. Embryologist burnout: physical and psychological symptoms and occupational challenges currently reported by US embryologists. *Fertility and Sterility* 118 (4), e66.
- Niederberger, C., Pellicer, A., Cohen, J., Gardner, D.K., Palermo, G.D., O'Neill, C.L., Chow, S., Rosenwaks, Z., Cobo, A., Swain, J.E., Schoolcraft, W.B., 2018. Forty years of IVF. *Fertility and sterility* 110 (2), 185–324.
- Ottolini, C., Rienzi, L., Capalbo, A., 2014. A cautionary note against embryo aneuploidy risk assessment using time-lapse imaging. *Reproductive biomedicine online* 28 (3), 273–275.
- Pacchiarotti, A., Selman, H., Valeri, C., Napolitano, S., Sbracia, M., Antonini, G., Biagiotti, G., Pacchiarotti, A., 2016. Ovarian stimulation protocol in IVF: an up-to-date review of the literature. *Current pharmaceutical biotechnology* 17 (4), 303–315.
- Palermo, G., Joris, H., Devroey, P., VanSteirteghem, A.C., 1992. Pregnancies after intracytoplasmic injection of single spermatozoon into an oocyte. *Lancet (London, England)* 340, 17–18. [https://doi.org/10.1016/0140-6736\(92\)92425-f](https://doi.org/10.1016/0140-6736(92)92425-f).

- Palmer, G.A., Paredes, O., Drakeley, A., Chavez-Badiola, A., Woolley, T.E., Kaouri, K., Cohen, J., 2024. Use and understanding of AI in the ART laboratory: an international survey. *Reproductive BioMedicine Online*, 104435.
- Pabon, Jr, J.E., Findley, W.E., Gibbons, W.E., 1989. The toxic effect of short exposures to the atmospheric oxygen concentration on early mouse embryonic development. *Fertility and Sterility* 51 (5), 896–900.
- Park, J.J., Tiefenbach, J., Demetriades, A.K., 2022. The role of artificial intelligence in surgical simulation. *Frontiers in Medical Technology* 4, 1076755.
- Pickering, S.J., Braude, P.R., Johnson, M.H., Cant, A., Currie, J., 1990. Transient cooling to room temperature can cause irreversible disruption of the meiotic spindle in the human oocyte. *Fertility and sterility* 54 (1), 102–108.
- Pribenszky, C., Nilselid, A.M., Montag, M., 2017. Time-lapse culture with morphokinetic embryo selection improves pregnancy and live birth chances and reduces early pregnancy loss: a meta-analysis. *Reproductive biomedicine online* 35 (5), 511–520.
- Puga-Torres, T., Blum-Rojas, X., Blum-Narváez, M., 2017 Jul. Blastocyst classification systems used in Latin America: is a consensus possible? *JBRA Assisted Reproduction* 21 (3), 222.
- Richardson, A., Brearley, S., Ahitan, S., Chamberlain, S., Davey, T., Zujovic, L., Hopkisson, J., Campbell, B., Raine-Fenning, N., 2015. A clinically useful simplified blastocyst grading system. *Reproductive biomedicine online* 31 (4), 523–530.
- Rienzi, L., Fauser, B., 2021. Future challenges for clinical embryologists. *Reproductive BioMedicine Online* 43 (6), 973–975.
- RStudio, R.T., Integrated development environment for R, 2020. RStudio, PBC: Boston, MA, USA. URL: <http://www.rstudio.com/>.
- Saiz, I.C., Gatell, M.C., Vargas, M.C., Mendive, A.D., Enedáguila, N.R., Solanes, M.M., Canal, B.C., López, J.T., Bonet, A.B., de, Mendoza, Acosta, M.V., 2018. The Embryology Interest Group: updating ASEBIR's morphological scoring system for early embryos, morulae and blastocysts. *Medicina Reproductiva y Embriología Clínica* 5 (1), 42–54.
- Saleh, F.L., Adashi, E.Y., Sable, D.B., Kushnir, V.A., Taylor, H.S., 2023. Changes to reproductive endocrinology and infertility practice, research, and training as investor mergers increase. *F&S Reports* 4 (4), 332–336.
- Sakkas, D., Shoukir, Y., Chardonnens, D., Bianchi, P.G., Campana, A., 1998. Early cleavage of human embryos to the two-cell stage after intracytoplasmic sperm injection as an indicator of embryo viability. *Human Reproduction (Oxford, England)* 13 (1), 182–187.
- Sawarkar, S., Griffin, D.K., Ribustello, L., Munné, S., 2021. Large intra-age group variation in chromosome abnormalities in human blastocysts. *DNA* 1, 91–104.
- Schoolcraft, W.B., Gardner, D.K., Lane, M., Schlenker, T., Hamilton, F., Meldrum, D.R., 1999. Blastocyst culture and transfer: analysis of results and parameters affecting outcome in two in vitro fertilization programs. *Fertility and sterility* 72 (4), 604–609.
- Sciorio, R., 2021 Apr. Use of time-lapse monitoring in medically assisted reproduction treatments: a mini-review. *Zygote* 29 (2), 93–101.
- Stephoe, P.C., Edwards, R.G., 1978. Birth after the reimplantation of a human embryo. *The Lancet* 312 (8085), 366.
- Storr, A., Venetis, C.A., Cooke, S., Susetio, D., Kilani, S., Ledger, W., 2015 Jul. Morphokinetic parameters using time-lapse technology and day 5 embryo quality: a prospective cohort study. *Journal of assisted reproduction and genetics* 32, 1151–1160.
- Tesarik, J., Greco, E., 1999. The probability of abnormal preimplantation development can be predicted by a single static observation on pronuclear stage morphology. *Human reproduction* 14 (5), 1318–1323.
- Tiegs, A.W., Tao, X., Zhan, Y., Whitehead, C., Kim, J., Hanson, B., Osman, E., Kim, T.J., Patounakis, G., Gutmann, J., Castelbaum, A., 2021. A multicenter, prospective, blinded, nonselection study evaluating the predictive value of an aneuploid diagnosis using a targeted next-generation sequencing–based preimplantation genetic testing for aneuploidy assay and impact of biopsy. *Fertility and sterility* 115 (3), 627–637.
- Tucci, V., Saary, J., Doyle, T.E., 2022. Factors influencing trust in medical artificial intelligence for healthcare professionals: A narrative review. *Journal of Medical Artificial Intelligence* 5.
- VerMilyea, M., Hall, J.M., Diakiw, S.M., Johnston, A., Nguyen, T., Perugini, D., Miller, A., Picou, A., Murphy, A.P., Perugini, M., 2020. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Human Reproduction* 35 (4), 770–784.
- VerMilyea, M., Hall, J.M., Diakiw, S., Johnston, A., Nguyen, T.D., Lim, A., Quangkananurug, W., Perugini, D., Murphy, A.P., Perugini, M., 2020. Camera-agnostic self-annotating artificial intelligence (AI) system for blastocyst evaluation. *ESHRE Virtual 36th Annual Meeting*.
- Wang, L., Wang, X., Liu, Y., Ou, X., Li, M., Chen, L., Shao, X., Quan, S., Duan, J., He, W., Shen, H., 2021 Dec. IVF embryo choices and pregnancy outcomes. *Prenatal Diagnosis* 41 (13), 1709–1717.
- Yang, M., Rito, T., Metzger, J., Naftaly, J., Soman, R., Hu, J., Albertini, D.F., Barad, D.H., Brivanlou, A.H., Gleicher, N., 2021 Apr. Depletion of aneuploid cells in human embryos and gastruloids. *Nature cell biology* 23 (4), 314–321.
- Yakar, D., Ongena, Y.P., Kwee, T.C., Haan, M., 2022. Do people favor artificial intelligence over physicians? A survey among the general population and their view on artificial intelligence in medicine. *Value in Health* 25 (3), 374–381.
- Zhan, Q., Sierra, E.T., 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 Reports* 1 (2), 133–141.
- Zhao, M., Li, H., Li, R., Li, Y., Luo, X., Li, T.C., Lee, T.L., Wang, W.J., Chan, D.Y., 2021. Automated and precise recognition of human zygote cytoplasm: A robust image-segmentation system based on a convolutional neural network. *Biomedical Signal Processing and Control* 67, 102551.

Received 20 December 2024; received in revised form 21 February 2025; accepted 28 March 2025.