



# Quality assurance (QA) for monitoring the performance of assisted reproductive technology (ART) staff using artificial intelligence (AI)

Panagiotis Cherouveim<sup>1</sup> · Victoria S. Jiang<sup>1</sup> · Manoj Kumar Kanakasabapathy<sup>2</sup> · Prudhvi Thirumalaraju<sup>2</sup> · Irene Souter<sup>1</sup> · Irene Dimitriadis<sup>1</sup> · Charles L. Bormann<sup>1</sup> · Hadi Shafiee<sup>2</sup>

Received: 16 June 2022 / Accepted: 20 October 2022 / Published online: 14 November 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

**Purpose** Deep learning neural networks have been used to predict the developmental fate and implantation potential of embryos with high accuracy. Such networks have been used as an assistive quality assurance (QA) tool to identify perturbations in the embryo culture environment which may impact clinical outcomes. The present study aimed to evaluate the utility of an AI-QA tool to consistently monitor ART staff performance (MD and embryologist) in embryo transfer (ET), embryo vitrification (EV), embryo warming (EW), and trophoctoderm biopsy (TBx).

**Methods** Pregnancy outcomes from groups of 20 consecutive elective single day 5 blastocyst transfers were evaluated for the following procedures: MD performed ET ( $N=160$  transfers), embryologist performed ET ( $N=160$  transfers), embryologist performed EV ( $N=160$  vitrification procedures), embryologist performed EW ( $N=160$  warming procedures), and embryologist performed TBx ( $N=120$  biopsies). AI-generated implantation probabilities for the same embryo cohorts were estimated, as were mean AI-predicted and actual implantation rates for each provider and compared using Wilcoxon signed-rank test.

**Results** Actual implantation rates following ET performed by one MD provider: “H” was significantly lower than AI-predicted (20% vs. 61%,  $p=0.001$ ). Similar results were observed for one embryologist, “H” (30% vs. 60%,  $p=0.011$ ). Embryos thawed by embryologist “H” had lower implantation rates compared to AI prediction (25% vs. 60%,  $p=0.004$ ). There were no significant differences between actual and AI-predicted implantation rates for EV, TBx, or for the rest of the clinical staff performing ET or EW.

**Conclusions** AI-based QA tools could provide accurate, reproducible, and efficient staff performance monitoring in an ART practice.

**Keywords** Artificial intelligence · Quality assurance · Trophoctoderm biopsy · Competency · Embryo transfer · Embryo vitrification and warming

## Introduction

Since the birth of the first baby conceived with assisted reproductive technology (ART) in 1978 [1], the field of ART has expanded exponentially with more than 200,000

ART procedures reported annually in the USA [2] and over 2 million procedures reported annually worldwide [3]. The penultimate goal of achieving a healthy live singleton birth relies on a multitude of factors, ranging from patient to physician to clinical contributors. As the volume and complexity of cases continue to increase, achieving and maintaining optimal staff performance in the ART practice is key [4, 5]. From an embryology perspective, successful fertilization, extended culture, and transfer depend on each individual embryologist’s technical dexterity, making them a crucial part of the ART process [4]. Despite their critical role to ART success, ART staff performance is innately “human” [6] with a high degree of variability between different embryologists [4, 5], making quality assurance/quality control among ART staff essential [7]. Timely identification

✉ Hadi Shafiee  
hshafiee@bwh.harvard.edu

<sup>1</sup> Division of Reproductive Endocrinology and Infertility, Vincent Obstetrics and Gynecology, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Suite 10A, Boston, MA 02114, USA

<sup>2</sup> Division of Engineering in Medicine, Brigham and Women’s Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA 02139, USA

of performance issues and variations becomes the biggest obstacle to overcome to achieve consistent outcomes and ensure the high-quality fertility care.

Quality assurance (QA) measures are key to reliably providing high-quality care over time and among different practices [8]. Accreditation bodies, such as the Clinical Laboratory Improvement Amendments of 1988 (CLIA) and the College of American Pathologists (CAP) recognize the need for continual quality monitoring and require every ART practice to continuously monitor for maintenance of accreditation [9]. Studies suggest that healthcare provider performance is one of the main determinants for successful embryo transfer (ET) [5, 10], prompting the Maribor [11] and the Vienna consensus [12] to propose key performance indicators (KPIs) within the ART setting. As implantation rate is considered to be a more reliable indicator of ART practice performance [12], embryo quality needs to be factored in as it has a significant impact on ART outcomes [13]. While grading was standardized by the Gardner's criteria [14], a significant degree of subjectivity, variability, and time [15, 16] exists with traditional manual grading, necessitating national semiannual proficiency tests among embryologists to ensure homogeneous embryo grading [17]. This lack of objectivity and consistency resulted in reliance on other indicators that take longer to detect, like clinical pregnancy, or are more rare, like no blastocyst survival after cryopreservation [11, 12]. Finally, embryo survival after trophoctoderm biopsy, embryo cryopreservation, or warming does not necessarily imply good implantation potential and pregnancy outcome [18]. Even though registering all supplies, including culture media, disposables, and other equipment, could be standardized and is regularly done to identify supply chain issues [7], efficient and reliable monitoring of staff performance remains a challenge.

Artificial intelligence (AI) serves as a cutting-edge tool to address this QA challenge. Although recently developed, eliminating subjectivity and ensuring high reproducibility have already made AI systems highly promising [19, 20] in the field of ART. AI-based image analysis networks, such as convolutional neural networks (CNN), integrate easily into fertility care [21], among many different medical specialties [22–25], as imaging plays a critical role in decision-making. AI's pattern recognition capabilities are far beyond human visual discrimination abilities [26]. CNNs have revolutionized computer vision with its ability to achieve exceptional consistency while removing human subjectivity from embryo grading [19, 20]. Also, emerging evidence supports AI's efficacy in predicting an embryo's developmental [27, 28] and implantation outcomes [29]. As such, in this study, we aimed to utilize AI-based tools as QA measures to assess the performance of attending physician (MD) and embryologist providers in a variety of procedures, including ET, embryo vitrification (EV), embryo warming (EW),

and trophoctoderm biopsy (TBx). AI models were used to predict implantation potential of the embryos and these predictions were compared with actual implantation outcomes. This way, we aimed to objectively factored in embryo quality to isolate and assess provider performance.

## Materials and methods

### Study design and embryo imaging

Data from ET performed at a single academic fertility center in Boston, Massachusetts, were retrospectively reviewed. This study was approved by Partners Healthcare Institutional Review Board (IRB# 2019P001000). Embryo development was recorded using EmbryoScope (Vitrolife, Sweden) time-lapse imaging. A Leica 20x -based imaging system with light from a single 635 nm LED was used to capture images every 10 min. Images included in this study were all acquired at specific time of embryo development (113 hours post insemination, hpi). All images were assessed by embryologists with at least 5 years of experience. The image collection was cleaned by removing completely indiscernible images of embryos. Embryos included in this study had images available at 113 hpi and known implantation outcomes. Embryos included in this study consisted of both untested and PGT-A tested euploid embryos comprising all cohorts.

### Artificial intelligence/convolutional neural network

The AI algorithm used for embryo assessments in this study was a CNN that was developed and trained as previously described [20, 30, 31]. Briefly, the previously developed CNN used the Xception architecture and was trained as a supervised binary classifier to predict implantation outcome using embryo images collected at 113 hpi. This AI system was around 75.3% accurate in predicting embryo implantation when assessing a cohort of euploid embryos [20, 30, 31].

### Embryo transfer, vitrification, warming, and trophoctoderm biopsy

Blastocyst quality higher than 3CC was considered as the criterion for EV and TBx [32]. Irvine Scientific Vitrification and Warming solutions (FujiFilm) was used for all EV and EW procedures. Embryos undergoing PGT were hatched on day 3 with a 1.48  $\mu\text{m}$  diode laser. Approximately 3–5 trophoctoderm cells were removed using the same laser. Eight MD providers (A–H) performed day 5 blastocyst transfers in the same academic fertility practice. Eight embryologists (A–H) performed ET, EV, and EW, while six embryologists (A–F) performed TBx.

AI-predicted and actual implantation outcomes were registered for the most recent 20 procedures (ET, EV, EW, or TBx) performed by each provider (MD or embryologist) and compared to assess operator performance. Throughout the study, all authors were blinded to the identity of the MD providers and the embryologists, in order to minimize associated bias.

**Statistical analysis**

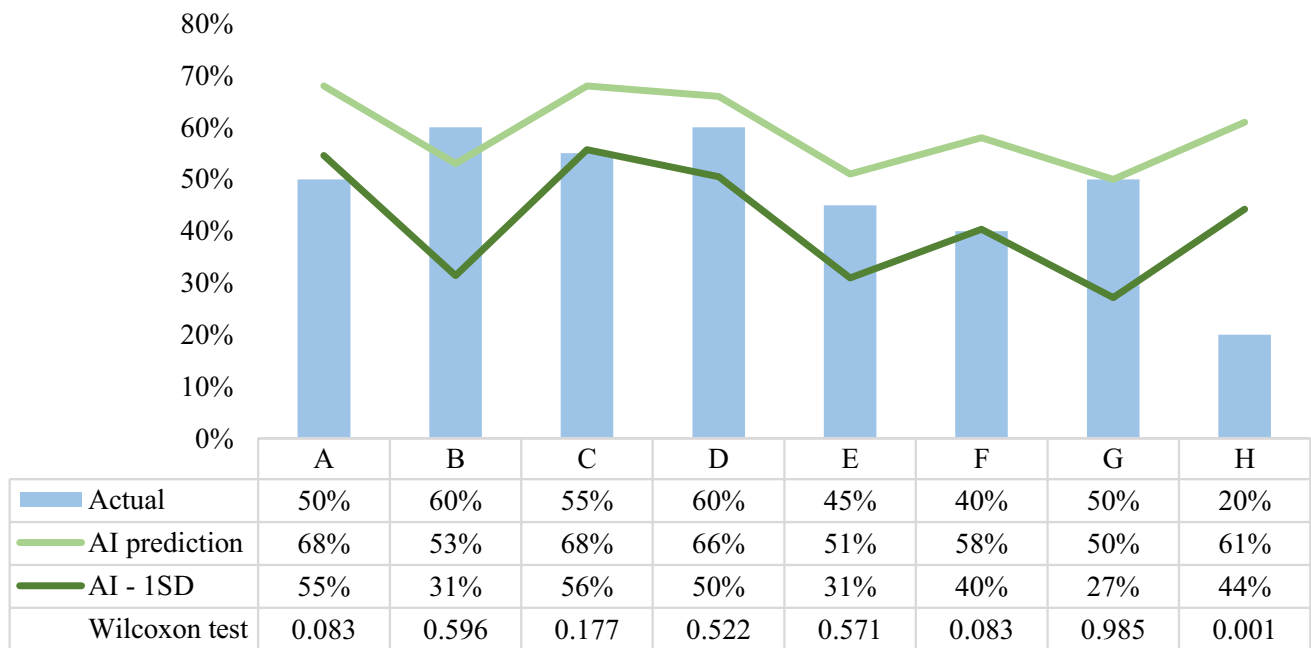
AI-predicted and actual implantation outcomes were registered for each embryo and classified according to provider (MD or embryologist) performing the procedure (ET, EV, EW, or TBx). Average AI-predicted, standard deviation (SD), and average actual implantation rates were estimated according to provider (MD or embryologist) and procedure performed (ET, EV, EW, or TBx). Wilcoxon signed-rank test was used to compare the AI-predicted and actual implantation outcomes of the same embryos within the same provider (MD or embryologist). Statistical significance was set as *p* value < 0.05. Statistical analysis was performed with R (v4.1.0; The R Foundation for Statistical Computing).

**Results**

**Embryo transfer performed by MD**

Implantation rates for 8 different MD providers (A–H) for their most recent 20 ETs were evaluated. Overall, in almost all cases (87.5%, 7/8 MD providers), actual implantation rates for ET performed by MD providers did not differ significantly (*p* > 0.05) from AI-based implantation prediction (Fig. 1). The greatest difference between the AI-predicted and the actual implantation rate was for MD provider H (61% vs. 20%, *p* = 0.001), which represented the only MD provider that had a performance significantly different from the AI-predicted performance. MD provider H’s implantation rate was also the lowest of the eight in our cohort (20% vs. 40–60%). The AI model prediction was perfectly accurate for MD provider G performance (AI prediction: 50% vs. actual: 50%, *p* = 0.985). Additionally, actual implantation rates following ET performed by most MD providers (62.5%, 5/8 MD providers) fell within 1 standard deviation (SD) of the AI prediction [AI prediction (SD) vs. actual: MD B, 53% (22%) vs. 60%; MD D, 66% (16%) vs. 60%; MD E, 51% (20%) vs. 45%; MD F, 58% (18%) vs. 40%; MD G, 50% (23%) vs. 50%] (Fig. 1).

**AI-predicted and actual implantation rates following Embryo Transfer performed by different MD providers**



**Fig. 1** Artificial intelligence (AI) predicted and actual implantation rates following embryo transfers (ET) performed by different MD providers

### Embryo transfer performed by embryologist

Implantation rates for 8 different embryologists (A–H) for their most recent 20 ETs were evaluated. AI prediction was not significantly different ( $p > 0.05$ ) from the actual implantation rates of the same embryos for all but one embryologist (87.5%, 7/8 embryologists) (Fig. 2). AI prediction was significantly different from the actual implantation rate in the case of embryologist H (AI prediction: 60% vs. actual: 30%,  $p = 0.011$ ). Embryologist H’s implantation rates were also the lowest observed in our embryologists’ cohort (30% vs. 50–70%) and the only one with more than 1 SD difference compared to AI prediction [AI prediction (SD) vs. actual: embryologist A, 63% (16%) vs. 60%; embryologist B, 57% (23%) vs. 65%; embryologist C, 62% (19%) vs. 70%; embryologist D, 62% (18%) vs. 65%; embryologist E, 64% (21%) vs. 50%; embryologist F, 62% (17%) vs. 65%; embryologist G, 63% (18%) vs. 60%] (Fig. 2). Finally, AI model’s average implantation predictions exhibited less variation in ET performed by embryologists compared to MD providers (57–64% vs. 50–68%).

### Embryo vitrification

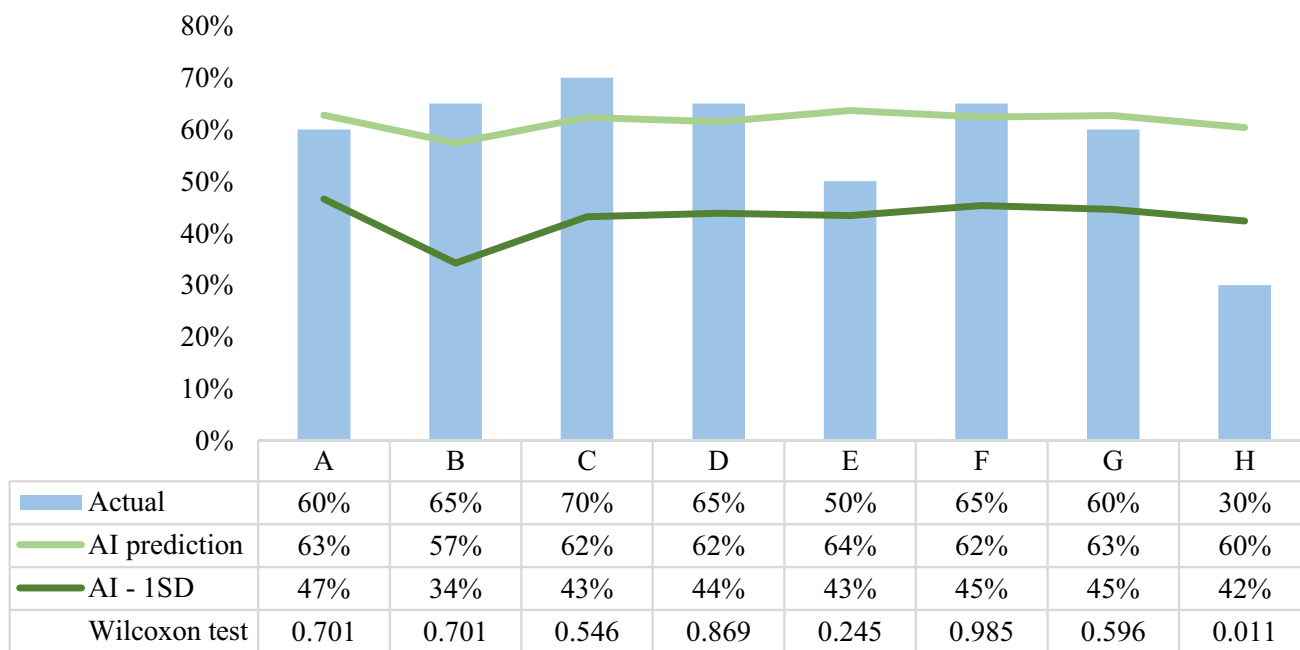
Regarding embryo vitrification performance, implantation rates of the 20 most recently vitrified embryos of each of the

8 embryologists (A–H) were compared with AI prediction. AI-predicted implantation rates did not differ significantly ( $p > 0.05$ ) from the actual rates observed following EV for any of the embryologists (100%, 8/8 embryologists) (Fig. 3). The largest disparity between AI prediction and actual implantation rate was following EV performed by embryologist G (AI prediction 63% vs. actual 45%), although it did not reach statistical significance ( $p = 0.076$ ). Implantation rates were closely predicted by AI (within 1 SD) among all embryologists except embryologist G (87.5%, 7/8) performing EV [AI prediction (SD) vs. actual: embryologist A, 56% (15%) vs. 65%; embryologist B, 65% (21%) vs. 60%; embryologist C, 63% (20%) vs. 65%; embryologist D, 52% (13%) vs. 55%; embryologist E, 56% (17%) vs. 55%; embryologist F, 61% (18%) vs. 55%; embryologist H, 57% (21%) vs. 60%] (Fig. 3).

### Embryo warming

Embryologists’ performance in warming embryos for transfer was assessed by comparing actual implantation rates of the 20 most recently warmed embryos of the 8 embryologists (A–H) with the rates predicted by the AI models. For most embryologists performing EW (87.5%, 7/8), actual implantation rates were not significantly different ( $p > 0.05$ ) from the AI-predicted implantation rate (Fig. 4). AI-prediction for

## AI-predicted and actual implantation rates following Embryo Transfer performed by different Embryologists



**Fig. 2** Artificial intelligence (AI) predicted and actual implantation rates following embryo transfers (ET) performed by different embryologists

### AI-predicted and actual implantation rates following Embryo Vitrification performed by different Embryologists

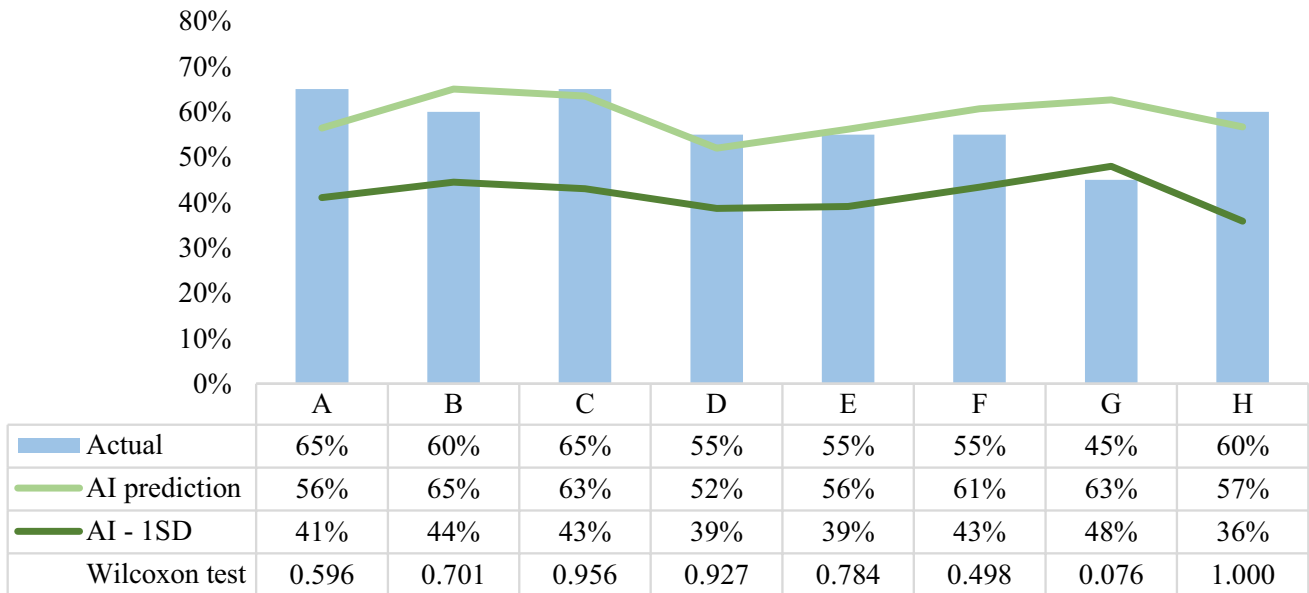


Fig. 3 Artificial intelligence (AI) predicted and actual implantation rates following embryo vitrification (EV) performed by different embryologists

### AI-predicted and actual implantation rates following Embryo Warming performed by different Embryologists

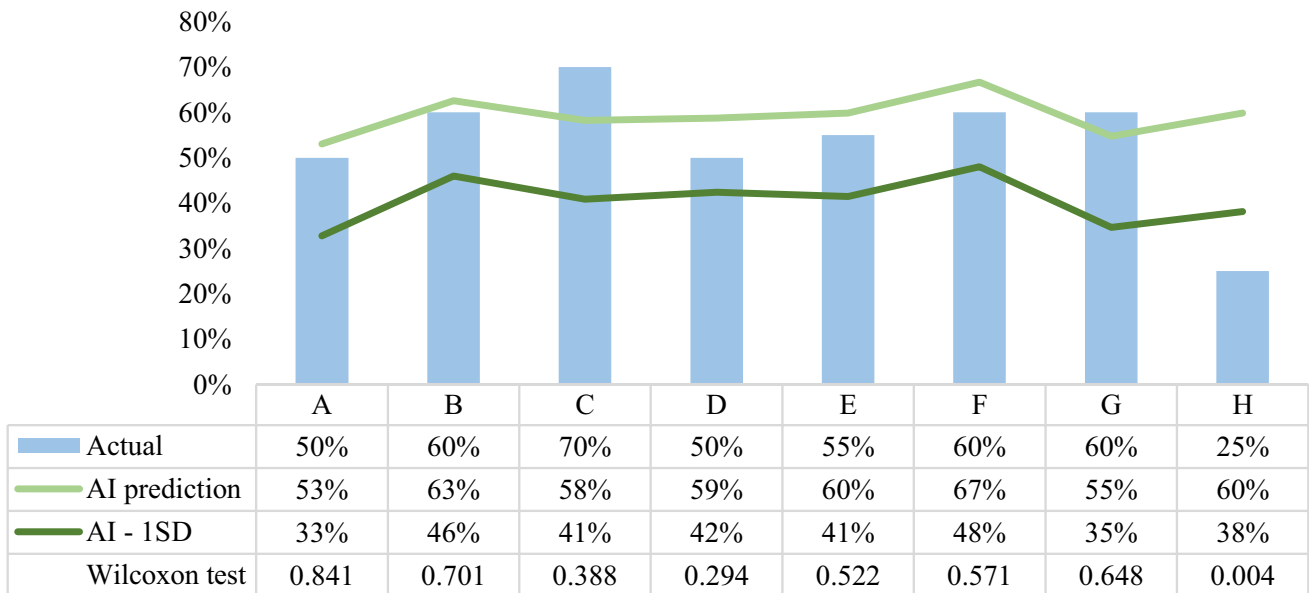


Fig. 4 Artificial intelligence (AI) predicted and actual implantation rates following embryo warming (EW) performed by different embryologists

implantation rate was significantly different from the actual only for embryos warmed by embryologist H (AI-predicted 60% vs. actual 25%,  $p=0.004$ ). Actual implantation rates for embryos warmed by embryologist H were the lowest observed in our embryologists' cohort (25% vs. 50–70%). All other embryologists (87.5%, 7/8) actual implantation rates were within 1 SD of the AI predictions [AI-predicted (SD) vs. actual: embryologist A, 53% (20%) vs. 50%; embryologist B, 63% (17%) vs. 60%; embryologist C, 58% (17%) vs. 70%; embryologist D, 59% (17%) vs. 50%; embryologist E, 60% (19%) vs. 55%; embryologist F, 67% (19%) vs. 60%; embryologist G, 55% (20%) vs. 60%] (Fig. 4).

### Trophectoderm biopsy

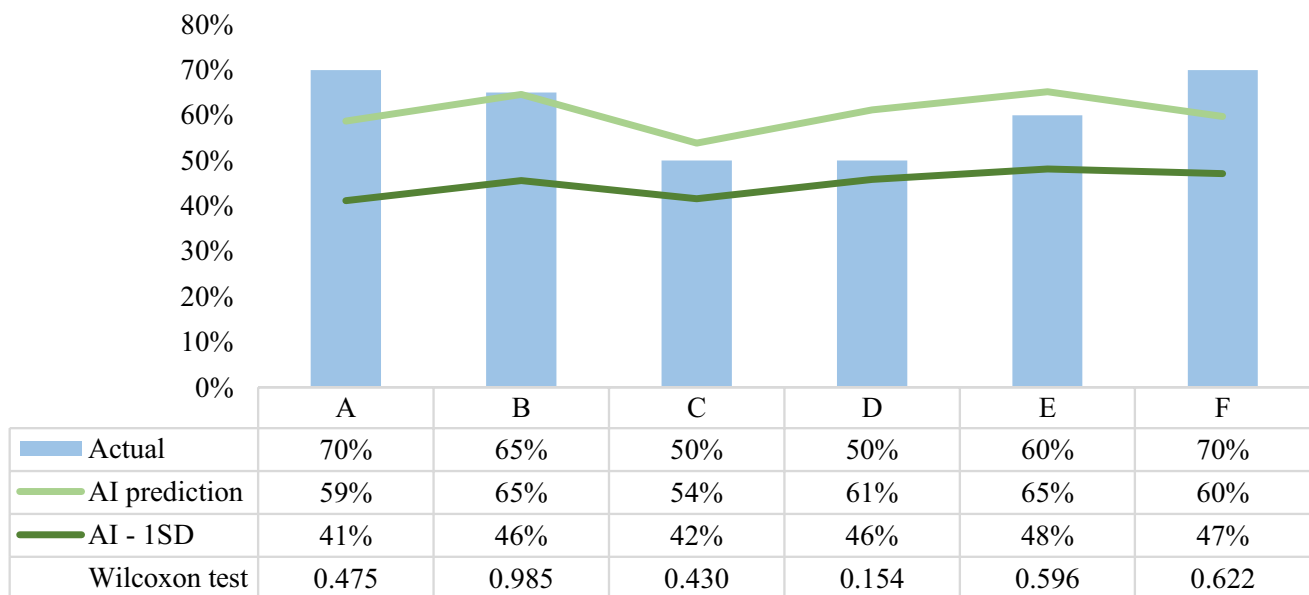
TBx performance of 6 different embryologists (A–F) was assessed by comparing AI-predicted with actual implantation rates of the 20 most recently biopsied embryos by each embryologist. AI-predicted implantation rates were not significantly different for the actual implantation rates of the cohort of embryos biopsied by all embryologists (100%, 6/6 embryologists). Additionally, AI predictions were within 1 SD of the actual implantation rates for embryos biopsied by all embryologists (100%, 6/6) [AI-predicted (SD) vs. actual: embryologist A, 59% (18%) vs. 70%; embryologist B, 65% (19%) vs. 65%; embryologist C, 54% (12%) vs. 50%;

embryologist D, 61% (15%) vs. 50%; embryologist E, 65% (17%) vs. 60%; embryologist F, 60% (13%) vs. 70%] (Fig. 5). The largest difference observed between AI prediction and actual implantation rates was in embryos biopsied by embryologist D (AI-predicted 61% vs. actual 50%) but was not statistically significant ( $p=0.154$ ) and within 1 SD of AI prediction. Among embryos biopsied by embryologist B, AI models were able to perfectly predict implantation rates (AI-predicted 65% vs. actual 65%).

### Discussion

Our study evaluated the utility of AI-predicted implantation rates as a QA measure for provider performance in a variety of procedures (ET, EV, EW, and TBx). Overall, results suggest that AI predictions were not significantly different and within 1 SD from actual implantation rates following procedures performed by most providers. Even though we found significant differences between predicted and actual implantation rates for one provider in our cohort performing ET and EW respectively, there were no significant differences for providers performing EV or TBx. Additionally, in the cases that there was a significant difference of more than 1 SD, actual implantation rates were lower than AI-predicted implantation rates, with that provider noted in having the

## AI-predicted and actual implantation rates following Embryo Biopsy performed by different Embryologists



**Fig. 5** Artificial intelligence (AI) predicted and actual implantation rates following embryo biopsy (TBx) performed by different embryologists



lowest implantation rates compared to the rest. In a different perspective, the gross deviations from the AI SD show the value of this system to objectively identify performance metrics that deviate from the clinic average. The AI system was able to identify MD A and H for ET and Embryologist H for ET and EW as outliers in performance, showing that this system can not only distinguish between providers but also integrate predicted implantation potential to the expected performance results. While all high-quality blastocysts have implantation potential, each embryo's implantation potential is not equal and needs to be individually assessed when using implantation potential as a KPI. For instance, for MD A within the ET evaluation, as the actual implantation rate as noted to be 50%, MD A may not have been marked during performance evaluation through traditional metrics evaluation. However, when including the AI-predicted implantation potential of the cohort, the MD A performance was noted to be more than 1 SD from the predicted potential, flagging their performance within the AI system. The AI prediction can serve as a benchmark of performance, with significant deviation potentially suggesting need for improvement in provider technique, serving as a warning signal for potential adverse quality performance events. With the aid of AI, QA tools could be standardized and would allow real-time adjustments to act upon and resolve quality performance issues that could impact an ART practice.

While AI system-generated implantation potential can be utilized as a reliable QA measure, the impact of this system is rooted in thoughtful integration and value-based implementation within a clinical system. A cutoff of 20 procedures was considered appropriate to assess performance over set time intervals, signifying appropriate safety checkpoints for continuous monitoring and adjustment of performance. This number of procedures in our practice roughly represents a monthly estimate of each provider's procedure workload. Specifying intervals for QA evaluation is important to ensure consistency through time and to promptly identify clinical deviations or procedural variations among providers to allow for timely intervention. We found that even this relatively limited number of procedures could make a major contribution to QA monitoring with the aid of AI. By utilizing these suggested AI-based QA measures, we could isolate the impact of operator performance from embryo quality on procedure outcomes and have access to more meaningful QA indicators than survival or degeneration rates. The number of procedures between quality assessment intervals could be adjusted according to individual practices and workload.

It has been well-described that ART practice success relies heavily on healthcare provider's performance [4, 5]. Cirillo et al. [5] found significant variation ET success among 32 operators performing ET, varying as much as 27% between individual operators. These observed differences were responsible for 44.5% of the heterogeneity in

ongoing pregnancy rates [5]. Interestingly, outcomes did not improve with an increased number of procedures performed by the provider [5], suggesting that the difference may originate from subtle differences among ET techniques utilized. Damage to the endometrial lining, uterine contractions induction, and depth of catheter insertion are integral parts of the ET technique and could all influence the outcome [10, 33, 34]. Similar concerns have been raised for embryologists' performance [4].

Two consensus meetings aimed to define clinical and laboratory quality indicators for ART practices [11, 12]. A key part of the quality management system was clearly defining and consistently monitoring performance indicators [35] to maintain the practice accreditation status [9]. Some of these indicators have proven their utility in timely detection of laboratory practices variations that could have a relevant clinical impact [36]. Regarding operator performance, an indicator suggested by the Maribor consensus to monitor ET competence is clinical pregnancy rate [11]. Clinical pregnancy, however, is a delayed outcome measure that could take weeks to quantify, limiting its ability as a prompt warning tool. A plausible alternative suggested by the Vienna consensus would be measuring implantation rates [12]. Implantation, however, can be impacted by a number of factors, such as embryo quality [13], leaving it a relatively subjective measure. Although embryo grading relies on specific predetermined criteria [14], traditional manual morphologic assessment is challenging, inherently reliant on a skilled human eye's subjective evaluation, resulting in high inter- and intra-observer variability [15, 16].

AI systems have a huge potential to assist ART practices in addressing this challenge. AI-based assessment is a highly objective and consistent [19] and can aid in process automation and acceleration, reducing the amount of time and effort required by the embryologist. AI is capable of accurately predicting embryo quality and developmental potential [27, 37]. To date, there has only been one study published demonstrating the utility of AI as a QA measure [20], which focused on determining the utility of an early warning AI-based system for culture monitoring and intracytoplasmic sperm injection (ICSI) competence assessment. None of the traditional indicators were able to identify culture conditions variations, except for the one AI-generated indicator ( $R^2 = 0.9063$ ) [20]. When evaluating competency in ICSI, results suggest minor differences between manual and AI determined fertilization, blastocyst, and high-quality blastocyst development [20]. AI systems have also been shown to be able to predict implantation potential of embryos reliably and accurately [29].

Strengths of this study include the robustness of the AI algorithm and the large pool of providers assessed. We included both MD and embryologist providers, showing the

power of AI-based QA among all aspects of ART practice. Additionally, we demonstrated the reliability and consistency of AI when assessing a wide range of practitioners while also demonstrating the variability among practitioners at baseline. Furthermore all cycles were performed within one institution, allowing for tight control of stimulation protocols, culture conditions, and practice management. Weaknesses of our study include that all CNNs were developed based only on Embryoscope images, with no images from other platforms included in our training, validation, or test sets. Additionally, although we have a mid- to large-sized practice, including more providers, both embryologists and MDs, would increase the reliability of our findings. Also, we trained our AI systems with 2440 images. As we increase the images available for use in training and validation, the accuracy of our AI systems will increase, decreasing the risk of unintentional bias introduced into the AI system when making predictions, and expanding our models to different settings and populations. Moreover, even though our AI-based models were able to identify deviations in individual's performance with only 20 procedures, future studies could implement these processes over a larger number of procedures and over a longer time period. The identity of the MD providers and embryologists were not revealed, and thus we did not have available data on specific techniques used and were not able to make the relevant comparisons. Finally, our approach did not account for certain patient cycle and transfer characteristics that could also influence pregnancy rates, including age, sperm quality, oocyte yield, and the complexity of the transfer.

Future studies should incorporate a larger sample size for training, validation, and testing, and should account for cycle or stimulation characteristics to allow for more accurate predictions. Additionally, assessing different intervals of testing would be helpful in meaningful integration of this technology based on clinic volume. Numeric scores could be implemented to represent blastocyst grades to further objectify these processes [38]. Moreover, apart from embryo quality and operator performance, there are patient, cycle, and transfer factors that could be considered to further enhance indicators utility [39]. Looking ahead, the future of AI systems seems extremely bright with a tremendous potential to provide an important edge for QA monitoring in many parts of an ART practice.

## Conclusion

Our study found evidence that AI systems could be implemented as a QA tool for staff performance monitoring. This study demonstrates the novel first application of AI models for quality control and KPI assessment in the ART laboratory. AI-based indicators could provide an edge over traditional indicators as they could be more objective, consistent, and time-efficient. In the future, additional potential

confounders could be taken into account by AI to provide a more reliable and holistic evaluation of each provider's performance.

**Author contribution** All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Panagiotis Cherouveim, Charles Bormann, Victoria S. Jiang, Manoj Kumar Kanakasabapathy, and Prudhvi Thirumalaraju. The first draft of the manuscript was written by Panagiotis Cherouveim, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This work was partially supported by the Brigham Precision Medicine Developmental Award and Innovation Evergreen Fund (Brigham Precision Medicine Program, Brigham and Women's Hospital), Partners Innovation Discovery Grant (Partners Healthcare): IGNITE Award (Connors Center for Women's Health & Gender Biology, Brigham and Women's Hospital) and R01AI138800, R01EB033866, R33AI140489, and R61AI140489 (National Institute of Health).

## Declarations

**Ethics approval and consent to participate** Informed consent was obtained from each individual before participation. Study protocols were approved by the Institutional Review Board (IRB#2019P001000) at Massachusetts General Hospital and Brigham and Women's Hospital.

**Consent for publication** Not applicable.

**Conflict of interest** Authors Dr. Hadi Shafiee, Dr. Charles Bormann, Prudhvi Thirumalaraju, and Manoj Kumar Kanakasabapathy wish to disclose a patent, currently licensed by a commercial entity, on the use of AI for embryology (US11321831B2). The rest of the authors declare that they have no competing interests.

## References

1. Steptoe PC, Edwards RG. Birth after the reimplantation of a human embryo. *Lancet*. 1978;312:366.
2. Sunderam S, Kissin DM, Zhang Y, Jewett A, Boulet SL, Warner L, et al. Assisted reproductive technology surveillance - United States, 2018. *MMWR Surveill Summ*. 2022;71:1–19.
3. Chambers GM, Dyer S, Zegers-Hochschild F, de Mouzon J, Ishihara O, Banker M, et al. International Committee for Monitoring Assisted Reproductive Technologies world report: assisted reproductive technology, 2014†. *Hum Reprod*. 2021;36:2921–34.
4. Go KJ. "By the work, one knows the workman": the practice and profession of the embryologist and its translation to quality in the embryology laboratory. *Reprod Biomed Online*. 2015;31:449–58.
5. Cirillo F, Patrizio P, Baccini M, Morengi E, Ronchetti C, Cafaro L, et al. The human factor: does the operator performing the embryo transfer significantly impact the cycle outcome? *Hum Reprod*. 2020;35:275–82.
6. Institute of Medicine (US) Committee on Quality of Health Care in America. *To Err is human: building a safer health system* [Internet]. Kohn LT, Corrigan JM, Donaldson MS, editors. Washington (DC): National Academies Press (US); 2000 [cited 2022 May 8]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK225182/>.



7. Wikland M, Sjöblom C. The application of quality systems in ART programs. *Mol Cell Endocrinol.* 2000;166:3–7.
8. ESHRE Guideline Group on Good Practice in IVF Labs, De los Santos MJ, Apter S, Coticchio G, Debrock S, Lundin K, et al. Revised guidelines for good practice in IVF laboratories (2015). *Hum Reprod.* 2016;31:685–6.
9. Olofsson JI, Banker MR, Sjöblom LP. Quality management systems for your in vitro fertilization clinic's laboratory: why bother? *J Hum Reprod Sci.* 2013;6:3–8.
10. Coroleu B, Barri PN, Carreras O, Martínez F, Veiga A, Balasch J. The usefulness of ultrasound guidance in frozen-thawed embryo transfer: a prospective randomized clinical trial. *Hum Reprod.* 2002;17:2885–90.
11. ESHRE Clinic PI Working Group, Vlasisavljevic V, Apter S, Capalbo A, D'Angelo A, Gianaroli L, et al. The Maribor consensus: report of an expert meeting on the development of performance indicators for clinical practice in ART. *Hum Reprod Open.* 2021;2021:hoab022.
12. ESHRE Special Interest Group of Embryology and Alpha Scientists in Reproductive Medicine. Electronic address: coticchio.biogenesi@grupposandonato.it. The Vienna consensus: report of an expert meeting on the development of ART laboratory performance indicators. *Reprod Biomed Online.* 2017;35:494–510.
13. Coello A, Nohales M, Meseguer M, de Los Santos MJ, Remohí J, Cobo A. Prediction of embryo survival and live birth rates after cryotransfers of vitrified blastocysts. *Reprod Biomed Online.* 2021;42:881–91.
14. Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. *Hum Reprod.* 2011;26:1270–83.
15. Storr A, Venetis CA, Cooke S, Kilani S, Ledger W. Inter-observer and intra-observer agreement between embryologists during selection of a single Day 5 embryo for transfer: a multicenter study. *Hum Reprod.* 2017;32:307–14.
16. Baxter Bendus AE, Mayer JF, Shipley SK, Catherino WH. Interobserver and intraobserver variation in day 3 embryo grading. *Fertil Steril.* 2006;86:1608–15.
17. Kovačić B, Prados FJ, Plas C, Woodward BJ, Verheyen G, Ramos L, et al. ESHRE Clinical Embryologist certification: the first 10 years. *Hum Reprod Open.* 2020;2020:hoaa026.
18. Zheng W, Yang C, Yang S, Sun S, Mu M, Rao M, et al. Obstetric and neonatal outcomes of pregnancies resulting from preimplantation genetic testing: a systematic review and meta-analysis. *Hum Reprod Update.* 2021;27:989–1012.
19. Bormann CL, Thirumalaraju P, Kanakasabapathy MK, Kandula H, Souter I, Dimitriadis I, et al. Consistency and objectivity of automated embryo assessments using deep neural networks. *Fertil Steril.* 2020;113:781–787.e1.
20. Bormann CL, Curchoe CL, Thirumalaraju P, Kanakasabapathy MK, Gupta R, Pooniwal R, et al. Deep learning early warning system for embryo culture conditions and embryologist performance in the ART laboratory. *J Assist Reprod Genet.* 2021;38:1641–6.
21. Dimitriadis I, Zaninovic N, Badiola AC, Bormann CL. Artificial intelligence in the embryology laboratory: a review. *Reprod Biomed Online.* 2022;44:435–48.
22. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* 2016;57:5200–6.
23. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542:115–8.
24. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18:500–10.
25. Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajirasouliha I. Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images. *EBioMedicine.* 2018;27:317–28.
26. Shapiro LG, Stockman GC. Computer vision. Prentice hall New Jersey; 2001.
27. Bortoletto P, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwal R, Souter I, et al. Predicting blastocyst formation of day 3 embryos using a convolutional neural network (CNN): a machine learning approach. *Fertil Steril.* 2019;112:e272–3 (Elsevier).
28. Dimitriadis I, Christou G, Dickinson K, McLellan S, Brock M, Souter I, et al. Cohort embryo selection (CES): a quick and simple method for selecting cleavage stage embryos that will become high quality blastocysts (HQB). *Fertil Steril.* 2017;108:e162–3 (Elsevier).
29. Fitz VW, Kanakasabapathy MK, Thirumalaraju P, Kandula H, Ramirez LB, Boehnlein L, et al. Should there be an "AI" in TEAM? Embryologists selection of high implantation potential embryos improves with the aid of an artificial intelligence algorithm. *J Assist Reprod Genet.* 2021;38:2663–70.
30. Bormann CL, Kanakasabapathy MK, Thirumalaraju P, Gupta R, Pooniwal R, Kandula H, et al. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. *eLife.* 2020;9:e55301.
31. Kanakasabapathy M, Dimitriadis I, Thirumalaraju P, Bormann CL, Souter I, Hsu J, et al. An inexpensive, automated artificial intelligence (AI) system for human embryo morphology evaluation and transfer selection. *Fertil Steril.* 2019;111: e11.
32. Gardner DK, Surrey E, Minjarez D, Leitz A, Stevens J, Schoolcraft WB. Single blastocyst transfer: a prospective randomized trial. *Fertil Steril.* 2004;81:551–5.
33. Davar R, Poormoosavi SM, Mohseni F, Janati S. Effect of embryo transfer depth on IVF/ICSI outcomes: a randomized clinical trial. *Int J Reprod Biomed.* 2020;18:723–32.
34. Fanchin R, Righini C, Olivennes F, Taylor S, de Ziegler D, Frydman R. Uterine contractions at the time of embryo transfer alter pregnancy rates after in-vitro fertilization. *Hum Reprod.* 1998;13:1968–74.
35. Mortimer ST, Mortimer D. Quality and risk management in the IVF laboratory [Internet]. 2nd ed. Cambridge: Cambridge University Press; 2015 [cited 2022 May 20]. Available from: <http://ebooks.cambridge.org/ref/id/CBO9781139680936>.
36. Hammond ER, Morbeck DE. Tracking quality: can embryology key performance indicators be used to identify clinically relevant shifts in pregnancy rate? *Hum Reprod.* 2019;34:37–43.
37. Dimitriadis I, Bormann CL, Thirumalaraju P, Kanakasabapathy M, Gupta R, Pooniwal R, et al. Artificial intelligence-enabled system for embryo classification and selection based on image analysis. *Fertil Steril.* 2019;111: e21.
38. Zhan Q, Sierra ET, Malmsten J, Ye Z, Rosenwaks Z, Zaninovic N. Blastocyst score, a blastocyst quality ranking tool, is a predictor of blastocyst ploidy and implantation potential. *F S Rep.* 2020;1:133–41.
39. Zacà C, Coticchio G, Vigiliano V, Lagalla C, Nadalini M, Tarozzi N, et al. Fine-tuning IVF laboratory key performance indicators of the Vienna consensus according to female age. *J Assist Reprod Genet.* 2022;39:945–52.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.