Research Article

ARTIFICIAL INTELLIGENCE IN EMBRYO SELECTION WHAT EVERY REPRODUCTIVE SPECIALIST SHOULD KNOW

***NIA KAVTARADZE**, MD Candidate; [1,2] **NINO MUSERIDZE**, MD, PhD[1]*

[1] *Georgian-German Reproduction Center, Tbilisi, Georgia*

[2] *David Tvildiani Medical University, Tbilisi, Georgia*

ABSTRACT

Background: Artificial intelligence (AI) is rapidly becoming an important component of modern in vitro fertilization (IVF), offering opportunities to standardize embryo assessment, reduce subjectivity, and efficiently analyze static and time-lapse imaging data.

Methods: A focused narrative review synthesized 17 publications from 2021 through 2025 on AI-assisted embryo ranking, viability and euploidy prediction, workflow optimization, quality assurance, and implementation: 1 randomized clinical trial, 1 systematic review, 2 conference abstracts, and 13 other journal publications. Four named embryo-assessment systems were profiled. Separately, published aggregate GGRC data were synthesized to illustrate a practice-based application of AI in IVF laboratory quality management.

Results: ERICA, iDAScore, FiTTE, and IVFvision. ai provided automated embryo ranking or outcome prediction. Most supporting evidence was retrospective or simulation-based. In the only included multicenter randomized trial (n = 1,066), iDAScore did not establish noninferiority to standard morphology-based selection for clinical pregnancy; however, embryo assessment was approximately tenfold faster. The GGRC evidence demonstrated a complementary quality-management application rather than image-based embryo ranking: a KPI-based clinical pregnancy model was externally validated using 3,888 Georgian treatment cycles within a two-center dataset of 10,128 cycles, yielding a mean AUC of 0. 73. In Q2 2024, predicted and observed clinical pregnancy rates were 58. 9% and 59. 1%, and no statistically significant between-embryologist differences were detected. These findings support AI-assisted calibration and laboratory audit but do not demonstrate improved pregnancy or live-birth outcomes.

Conclusion: AI's strongest current contribution is human-supervised decision support and laboratory governance. The GGRC experience demonstrates practical value for calibration, KPI surveillance, troubleshooting, and targeted audits, but it does not yet prove that AI increases pregnancy or live-birth rates. Thoughtfully implemented AI can strengthen standardization,

CITATION: *Kavtaradze N, Museridze N. Artificial intelligence in embryo selection: What every reproductive specialist should know. Medical Times. Georgian-German Reproduction Center; 2026;(5):1-12. https://doi.org/10.71419/mtggrc.2026.36*

© 2026 Kavtaradze, Museridze. This article is published in Medical Times by the Georgian-German Reproduction Center under the Creative Commons Attribution 4.0 International License (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

workflow, training, quality control, and evidence-based IVF care. However, it should not replace embryologists or be presented as a guarantee of euploidy, pregnancy, or live birth.

Keywords: artificial intelligence; embryo selection; IVF; time-lapse imaging; machine learning; deep learning; euploidy; reproductive medicine

Introduction

Embryo selection is central to IVF because clinicians aim to identify the embryo most likely to implant while supporting safe single-embryo transfer.[1] Morphology remains the clinical benchmark: it is inexpensive, familiar, and biologically meaningful.[1] Yet it is also subjective, depends on observer experience, and is usually based on a limited number of static observations.[2]

AI became relevant as time-lapse incubators and digital imaging generated more data than can be reviewed efficiently by manual methods alone.[3,4] Machine-learning systems can apply the same scoring rules repeatedly and may recognize complex visual or developmental patterns that are difficult to detect consistently by human observers.[2-4] Figure 1 shows the main reasons for using AI in embryo assessment.

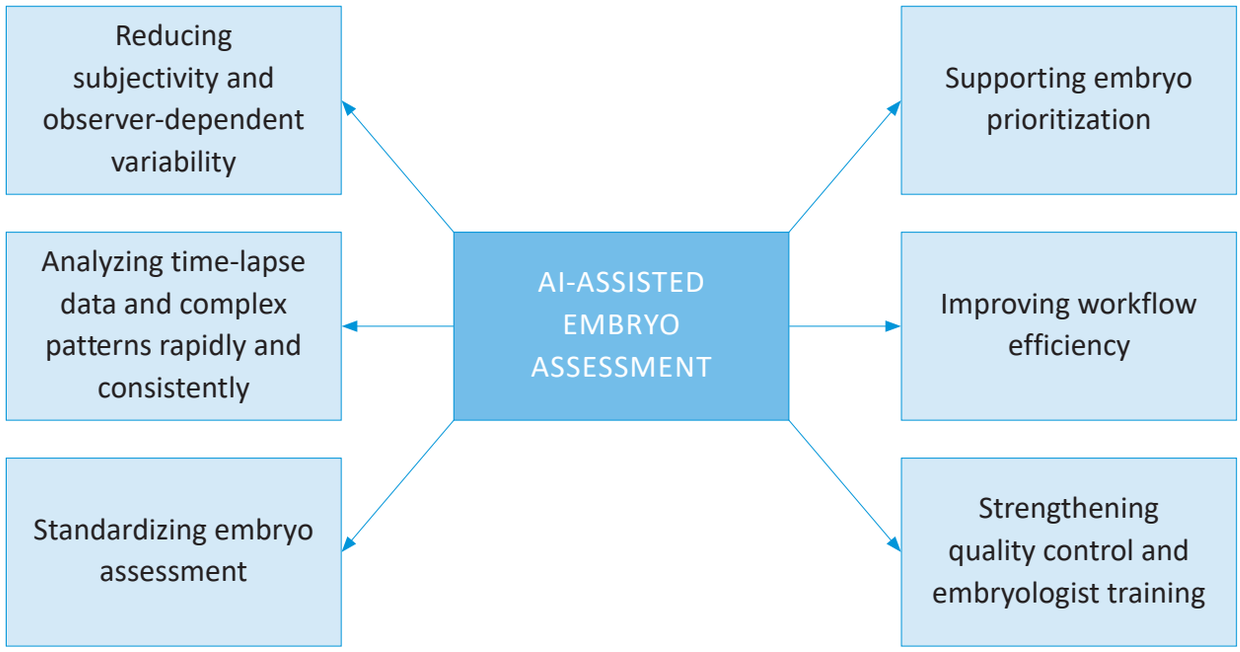

Figure 1. The clinical drivers for AI adoption in embryo assessment.

Three messages for reproductive specialists

- **Why AI is relevant:** IVF laboratories need greater consistency, faster review, and better use of image-rich data.
- **What AI can do today:** support embryo ranking, workflow optimization, standardization, training, and quality control.
- **What AI cannot yet do:** replace the embryologist, guarantee euploidy, or prove an increase in live birth across routine practice.

Methods

A focused narrative review was conducted in PubMed and Google Scholar for English-language publications on AI-assisted embryo ranking, viability and euploidy prediction, workflow optimi-

zation, quality assurance, and implementation, dated January 1, 2021, through December 31, 2025. Seventeen publications were included in the qualitative synthesis: 1 randomized clinical trial, 1 systematic review, 2 conference abstracts, and 13 other journal publications. Multicenter and multi-clinic evidence was retained within these design categories. Four named embryo-assessment systems – ERICA, iDAScore, FiTTE, and IVFvision. ai – were profiled. Separately, published aggregate GGRC data were synthesized to illustrate a practice-based application of AI in IVF laboratory quality management.

Search concepts combined "artificial intelligence" or "machine learning" with "embryo selection", "embryo ranking", "time-lapse", "blastocyst", "euploidy", "implantation", "clinical pregnancy", "IVF", or "laboratory KPI. " Eligible reports involved human IVF and provided a clinically or operationally interpretable outcome, including embryo ranking, viability or euploidy prediction, implantation or pregnancy prediction, assessment time, external validation, laboratory quality control, or staff-performance monitoring. Animal-only studies, publications outside the date or language limits, and purely technical studies without a clinically interpretable outcome were excluded. Conference abstracts were retained as preliminary evidence and interpreted cautiously.

Data extraction included study design, sample size, clinical setting, model inputs, comparator, target outcome, validation approach, performance metrics, and reported limitations. Evidence was weighted according to study design and validation strength. Randomized prospective evidence was given the greatest weight in assessing clinical effectiveness, while the systematic review was used to contextualize the broader evidence base. Multicenter studies and independent external validation informed the assessment of generalizability, whereas single-center retrospective studies and conference abstracts were regarded as preliminary evidence. Because the included studies differed substantially in model architecture, input data, comparators, outcome definitions, and clinical endpoints, quantitative pooling was considered inappropriate; findings were therefore synthesized narratively.

The practice-based GGRC component synthesized previously published aggregate data reported by Sergeev et al, including laboratory KPIs, model calibration over time, observed versus predicted clinical pregnancy rates, staff-level comparisons, discordant-case analysis, and documented quality-management actions.[11,16,17] No new patient-level analyses or statistical reanalyses were conducted. Because the published GGRC reports did not include a controlled preimplementation comparison, temporal changes were interpreted as changes in prediction-outcome agreement and model calibration – not as evidence that AI increased clinical pregnancy or live-birth rates or reduced variability between embryologists.[11]

Results

Why AI Became Necessary in Embryology

Conventional morphology captures important features, but embryo grading can vary among embryologists and even within the same observer over time.[1,2] This variability becomes more important in high-volume laboratories, multicenter networks, and situations in which several embryos appear morphologically similar.[2,5]

Time-lapse imaging introduced continuous monitoring of embryo development, allowing detailed assessment of key morphokinetic events, including pronuclei appearance and fading (tPNa and tPNf), cleavage timings (t2-t8), morula compaction (tM), the start of blastulation (tSB), blastocyst expansion, cell-cycle progression, and developmental synchronization.[3] These

recordings are clinically valuable but time-consuming to annotate manually.[3] Automated analysis can process thousands of images efficiently and convert complex developmental data into reproducible scores or standardized developmental profiles.[3,4]

Importantly, AI does not improve the intrinsic quality of an embryo.[1,5] Its role is to prioritize the embryos already available and to make assessment more consistent, efficient, and auditable.[1,4,5]

How AI Works – Without the Technical Jargon

Machine learning identifies relationships between defined inputs – such as morphology, cleavage timing, maternal age, or laboratory variables – and a known outcome.[2] Deep learning is a more automated form of machine learning that can learn visual patterns directly from embryo images or image sequences.[2,4] Static-image systems analyze one or a small number of selected images, whereas time-lapse systems analyze development across multiple time points.[2,4,6,7] Multimodal systems combine images with clinical or laboratory information.[6,8] Figure 2 illustrates the simplified workflow of AI-assisted embryo assessment. The output is usually a ranking or probability – for example, predicted viability, implantation potential, or likelihood of euploidy.[6,8] These outputs support embryo prioritization but should not be interpreted as a diagnosis or guarantee of pregnancy.[8,9]

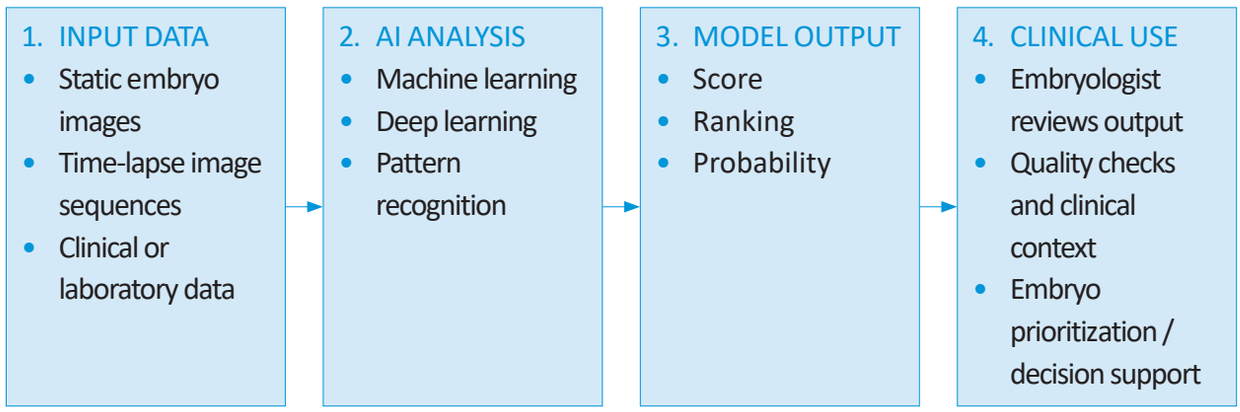

Figure 2. Main data inputs and outputs of AI-assisted embryo assessment models.

AI systems also differ in the transparency of their decision-making processes. Some models function as “black boxes”, producing predictions without clearly showing which features influenced the results.[2] More interpretable or hybrid systems allow embryologists to examine the morphological, morphokinetic, or clinical variables that contributed to a prediction.[2] Lee et al classified AI decision-support systems according to their degree of interpretability, as summarized in Table 1.[2] This distinction is clinically important because greater transparency may improve user trust, quality control, and the embryologists” ability to understand or question an AI-generated recommendation.

Table 1. Classification of AI Decision-Support Systems for Embryo Selection According to Interpretability.

Category	Annotation Step	Ranking Step	Interpretability
Black-box	None	Black-box methods	Lowest
Matte-box	Manual or automatic	Black-box methods	Moderate
Glass-box	Manual or automatic	Interpretable ML (decision tree, logistic regression, Bayesian network, random forest)	Highest

Adapted by the author from Lee et al.[2] Abbreviations: ML, machine learning.

Programs Currently Discussed in Clinical Embryology

The available systems differ in their input data, intended outcome, transparency, and evidence base.[2,4,6,7,10] They should not be treated as interchangeable products.[2,5] Importantly, not every clinically useful AI system directly ranks embryos. Four embryo-assessment systems are compared in Table 2. The GGRC model is separate because it combines laboratory and clinical KPIs for pregnancy prediction, audit, staff assessment, and quality improvement rather than ranking embryo images.[11]

What the Evidence Shows

Systematic reviews and retrospective studies generally show that AI can match or outperform individual embryologists on selected image classification or prediction tasks.[13] AI can markedly reduce annotation time and provide consistent scoring.[4,12] Automated morphokinetic analysis can identify developmental patterns associated with implantation-related outcomes.[3]

Promising discrimination does not automatically translate into improved live birth.[5,8,12,13] Models trained on morphology, implantation, biochemical pregnancy, or euploidy answer different questions. For example, a multimodal model predicted euploidy with good discrimination but did not effectively predict live birth after frozen embryo transfer. 8 Live birth reflects the entire reproductive and obstetric pathway and cannot be inferred from embryo appearance alone.[8,12,13]

The strongest prospective evidence remains cautious.[12] In a multicenter randomized trial of 1,066 patients, clinical pregnancy occurred in 46.5% of the iDAScore group and 48.2% of the morphology group.[12] The trial did not establish noninferiority using its prespecified margin.[12] However, embryo evaluation was approximately tenfold faster with AI.[12]

Table 2. Selected AI systems and their current clinical interpretation.

System	Primary input	Current purpose and evidence	Practical caution
ERICA	Single static blastocyst image	Ranks embryos using a noninvasive image.[7] In a 2025 controlled simulation study, ERICA outperformed participating embryologists in selecting a euploid embryo on the first attempt.[7] It can support embryo prioritization, decision-making, and embryologist training.[7]	Its effect on live-birth outcomes in routine practice has not been established.[5,7] Performance depends on image acquisition and workflow integration.[5,7]
iDAScore	Time-lapse image sequence	Provides fully automated embryo ranking by analyzing spatial and temporal patterns throughout embryo development.[4] A large retrospective multicenter study supported reproducible scoring and demonstrated partial generalizability across previously unseen clinics.[4] In a randomized trial, evaluation was about 10-fold faster than with standard morphology assessment.[12]	The randomized trial did not demonstrate noninferiority for clinical pregnancy compared with standard morphology-based selection.[12] Therefore, current evidence supports its value for automation, standardization, and workflow efficiency, but does not establish superior pregnancy or live-birth outcomes.[5,12]
FiTTE	Single static day-5 blastocyst image, optionally combined with clinical variables.	Predicts blastocyst viability and the probability of clinical pregnancy after single-embryo transfer.[6] Grad-CAM provides visual explanations of the image regions contributing to the prediction.[6] Performance improved when image data were combined with clinical variables, illustrating the potential of multimodal AI.[6]	Evidence is based on a retrospective, single-center study and requires broader external and prospective validation.[5,6] Visual explanations may improve interpretability but do not establish causality or eliminate bias and uncertainty.[6,9]
IVFvision.ai	Single static day-5 blastocyst image	Predicts implantation potential.[10] Preliminary external validation data suggest that image centering and the addition of clinical or laboratory variables, such as fertilization method and KIDScoreD5, can materially affect performance.[10]	Evidence is limited to a retrospective conference abstract based on a small sample of 113 blastocyst images.[10] Performance was highly dependent on image quality and blastocyst positioning, highlighting the need for standardized image acquisition and larger prospective validation studies.[10]

Table developed by the author based on published evidence.

Practice-Based GGRC Experience: From Clinical Pregnancy Prediction to IVF Quality Management

This practice-based synthesis describes a broader practical application of AI in IVF laboratory management, using published aggregate data from GGRC and related multicenter validation and audit reports; it is not a new patient-level analysis.[11,16,17] Unlike embryo-ranking tools, this system combines clinical variables and laboratory KPIs to estimate the expected probability of clinical pregnancy. Observed outcomes can then be compared with model-generated predictions for defined patient groups, time periods, and staff members to identify discrepancies requiring quality review.

Data foundation and external validation: The GGRC external validation dataset included 3,888 treatment cycles (2022-2024) and four embryologists. With a second external center, 10,128 total cycles yielded a mean AUC of 0.73.[16] A subsequent GGRC analysis included 665 treatment cycles with embryo-transfer outcomes from 2024.[17] Non-calibrated AUCs across nine other clinics ranged from 0.55 to 0.73, emphasizing the importance of validating and calibrating prediction models for each clinic"s patient population and clinical practices.[17]

Real laboratory KPIs: In the later GGRC analysis, mean fertilization, blastocyst-development, and total good-quality blastocyst development rates were 0.83 (95% CI, 0.70-0.96), 0.52 (0.48-0.56), and 0.43 (0.37-0.50), respectively, compatible with consensus targets.[17] In the externally validated prediction model, LIME-derived values associated with classification into the positive-pregnancy group included an oocyte-retrieval rate of 0.83, MII rate of 0.78, fertilization rate of 0.64, blastocyst-development rate of 0.44, TGBDR of 0.35, and KPIScore of 17.[16] These are interpretive thresholds for the analyzed datasets and should not be regarded as universal clinical cutoffs or causal requirements for pregnancy.

Model calibration over time: No controlled comparison of clinical outcomes before and after AI implementation was reported. Mean absolute error between predicted and observed aggregate pregnancy rates fell from 0.15 in 2022 Q1 to 0.03 in 2024 Q2 ($p < 0.05$) as data increased and the algorithm was refined. In Q2 2024, following integration of the quality-control system, the predicted clinical pregnancy rate was 58.9%, compared with an observed rate of 59.1%, representing a difference of 0.2 percentage points.[17] These findings demonstrate improved model calibration.

Embryologist consistency and targeted audit. Pairwise comparisons of laboratory KPIs among embryologists in Q2 2024 did not detect statistically significant differences ($p > 0.05$). Among 64 discordant cases identified within 665 transfers, differences associated with the embryologist were also nonsignificant ($p = 0.633$).[17] This supports current consistency. However, nonsignificant findings do not prove equivalence between embryologists, and without a pre-implementation baseline, they cannot demonstrate that variability decreased after AI implementation. Nevertheless, identifying discordant cases provides a practical targeted audit approach: instead of reviewing all cycles equally, the laboratory can focus on further investigating cases in which predicted and observed outcomes differ and assess potential procedural, clinical, data-quality, or model-calibration explanations.[17]

Concrete troubleshooting. A 10%-20% decline in 2023 pregnancy probability occurred without significant changes in laboratory KPIs (all $p > 0.05$), redirecting review toward clinical or patient factors. The largest model-outcome discrepancy coincided with 2022 Q2 laboratory reconstruction; that period was excluded.[11] Because this was a retrospective observational analysis, it cannot establish the cause of the decline.

Overall, the GGRC experience demonstrates how AI-based clinical-pregnancy prediction can extend beyond individual embryo assessment to support KPI surveillance, model calibration, temporal monitoring, staff-level audit, and structured investigation of unexpected outcomes. Nevertheless, these retrospective findings do not establish that AI independently improves pregnancy or live-birth rates or reduces variability between staff members. Its demonstrated value lies in providing an additional, data-driven component of a human-supervised IVF quality-management system. Table 3 summarizes the principal published GGRC findings and their practical interpretation within an IVF quality-management framework.

Where the Greatest Benefit Is Likely Today

- **Quality control:** calibrated predictions and KPI trends can distinguish a laboratory signal from changes in case mix or clinical practice.
- **Staff development:** case-mix-adjusted audit identified physicians whose observed outcomes were below predicted performance and supported focused supervision and mentoring.[11]
- **Standardization:** the same algorithm can apply a consistent scoring framework across shifts, operators, and affiliated laboratories.[2,4,11]
- **High-volume IVF centers:** automated review can reduce time spent on routine ranking and allow experts to focus on complex cases and laboratory oversight.[4,5,12]

Table 3. Published GGRC Metrics and Their Quality-Management Interpretation.

GGRC measure	Published result	Quality-management interpretation
External validation	3,888 GGRC protocols (2022-2024), evaluated by four embryologists; two-center external validation: 10,128 cycles, mean AUC 0.73.[16]	Supports model generalizability; local validation and calibration remain necessary.
Laboratory KPIs	Fertilization rate 0. 83; blastocyst development rate 0. 52; TGBDR 0.43.[11]	Provides real laboratory benchmarks for investigating changes in clinical pregnancy rate.
Model calibration	Quarterly mean absolute error decreased from 0.15 to 0.03 ($p < 0.05$); predicted vs observed Q2 2024 clinical pregnancy rates were 58.9% and 59.1%.[17]	Demonstrates closer agreement between predicted and observed rates over time; doesn”t show improvement in clinical outcomes.
Embryologist comparison	Pairwise KPI comparisons, $p > 0.05$; among 64 discordant cases from 665 embryo transfers, differences associated with the embryologist were nonsignificant ($p = 0. 633$). [17]	No statistically detectable between-embryologist differences were found in the measured analyses; no preimplementation baseline was available to assess whether variability changed.

Practice-based synthesis of published aggregate GGRC data and related validation analyses.[16,17] Abbreviations: AUC, area under the curve; CI, confidence interval; KPI, key performance indicator; TGBDR, total good-quality blastocyst development rate.

Limitations

Most published studies remain retrospective, and many are based on single-center or selected datasets.[5,13] Training data may underrepresent difficult embryos, untransferred embryos, certain patient groups, or specific laboratory conditions.[4,13] Overfitting can produce excellent performance in the development dataset but weaker performance in a new clinic.[4,5]

External validation is limited, and performance can change with incubator type, image acquisition, culture conditions, transfer policy, patient population, and endpoint definition.[4,5,10] Small technical issues matter: preliminary IVFvision. ai data showed substantially different performance for centered versus off-centered blastocyst images.[10]

Black-box algorithms create problems for trust, counseling, responsibility, and audit.[9,15] Bias is also possible.[14,15] A large conference abstract reported small sex-related scoring differences in both manual and some AI methods; the differences were not sufficient for reliable sex selection, but they emphasize the need for subgroup and fairness testing.[14]

Cost, software updates, data privacy, licensing, staff training, and performance drift must also be considered.[5,15] A model should be evaluated not only by accuracy but by whether it fits the laboratory workflow, remains calibrated over time, and produces a measurable clinical or operational benefit.[5,15]

Discussion

What the GGRC Data Add to Real-World IVF Practice

Most embryo-selection studies ask a narrow question: which embryo should be transferred first? The GGRC program addresses a different operational question: whether the complete clinical-laboratory pathway is performing as expected for defined patient groups and time periods.[11,16,17] This extends AI beyond single-embryo ranking to an auditable quality-management framework.

Three findings are particularly relevant. Firstly, stable laboratory KPIs during a decline in predicted clinical pregnancy probability help determine that deterioration in laboratory performance is an unlikely explanation. Secondly, closer agreement between predicted and observed pregnancy rates over time demonstrates the practical value of model calibration and continuous performance monitoring. Calibrated predictions reveal staff-level outliers that require supervision. And thirdly, nonsignificant measured differences among embryologists provide reassurance about current consistency. Together, these findings demonstrate practical value for monitoring, troubleshooting, and focused quality review.

A safe workflow begins with technical quality.[5,10] Poor focus, off-centered blastocyst images, incomplete time-lapse data, or image-acquisition conditions that differ from the model's training data can make an apparently precise score unreliable.[5,10] The embryologist should perform standard assessment, review the AI result and any available confidence or quality indicators, and investigate clinically important disagreement before making the final decision.[5,9]

Patients should be told that AI provides an estimate or ranking, not a guarantee.[5,9,15] Patient counseling should explain what data are used, whether the software is locally validated, and that a high score cannot confirm euploidy or ensure pregnancy and live birth.[5,8,9,15]

How Clinics Should Implement AI

- Define the intended use before purchase or development: embryo ranking, time-lapse annotation, clinical pregnancy prediction, KPI monitoring, quality control, staff competency assessment, training, or workflow efficiency.

- Record a pre-implementation baseline and validate the system locally against current laboratory practice before allowing it to influence embryo prioritization or staff assessment.
- Measure operational outcomes (assessment time, calibration error, observer variability, detected outliers, and completed corrective actions) alongside clinical outcomes (implantation, clinical pregnancy, miscarriage, and live birth).
- Link alerts and important AI-human disagreements to a documented corrective and preventive action pathway, including review, mentoring when indicated, and repeat audit.
- Reassess model performance after software updates or changes in incubators, laboratory protocols, or patient populations.

Table 4. Current Practical Value and Unproven Claims of AI in Embryo Assessment.

What current evidence supports	What remains unproven	Clinical interpretation
Standardized scoring: Automated systems apply the same scoring method consistently.[2,4]	Generalizability remains incompletely demonstrated: some models performed across unseen clinics, whereas many studies lacked external validation or showed site-dependent performance.[4,5,10,13]	Validate performance locally and monitor it over time.
Faster assessment: iDAScore evaluation was almost 10 times faster than manual morphology assessment.[12]	The randomized trial did not establish noninferiority for clinical pregnancy or show improved live birth.[12]	Workflow efficiency is a valid benefit, even without better clinical outcomes.
Euploidy estimation: A multimodal model combining embryo and clinical data showed good discrimination for euploidy.[8]	AI cannot confirm chromosomal status and did not reliably predict live birth after euploid embryo transfer.[8]	Treat the result as a probability estimate, not a diagnosis or substitute for PGT-A.
Embryo prioritization: iDAScore supported implantation-related ranking, while ERICA selected a euploid embryo first more often than embryologists in a simulated task.[4,7]	ERICA was evaluated in a simulation, not an embryo-transfer trial. After viewing ERICA rankings, embryologists" final selections did not improve significantly.[7]	Use AI as decision support, not as a replacement for expert assessment.
Laboratory audit and staff benchmarking: KPI-based analysis identified differences in staff performance across clinics and supported targeted review.[11]	Improved long-term competency, pregnancy rates, or live-birth outcomes after mentoring were not demonstrated.[11]	Use AI as a quality-management tool with predefined KPIs and follow-up evaluation.
Training and feedback: AI can provide a consistent benchmark for structured review of discordant cases.[7,11]	A sustained improvement in embryologist accuracy or competency has not been established.[7]	Training value is promising but should be seen as potential rather than proven.

Table developed by the author based on published evidence.

Future Priorities

Future research should prioritize prospective multicenter trials with live birth, cumulative live birth per oocyte retrieval, time to pregnancy, miscarriage, obstetric and neonatal outcomes, cost-effectiveness, and patient experience.[5,12,15]

Interpretability and fairness should be built into validation rather than treated as optional additions.[9,14,15] Multimodal systems may ultimately be more useful than image-only models, but they will require larger datasets, stronger privacy protections, calibration, and ongoing surveillance.[5,6,8,9,15]

Conclusion

Artificial intelligence is no longer only a future concept in embryology; it is already increasingly being evaluated and introduced into IVF practice.[5,11,12] Its greatest value today lies not in replacing embryologists, but in supporting standardized, reproducible, and evidence-based embryo assessment, workflow efficiency, and laboratory quality management.[2,5,11,12]

Across 17 included publications, the clearest benefits of AI are standardized scoring, faster assessment, structured embryo ranking, model calibration, and laboratory audit. Superiority for clinical pregnancy or live birth has not been established.[4,5,12,13,17] The GGRC program provides a concrete practice-based example through a validation cohort of 3,888 treatment cycles, center-specific laboratory KPIs, close agreement between predicted and observed pregnancy rates, and targeted analysis of discordant cases.[11,16,17] Responsible implementation requires human-supervised AI that is validated locally, monitored continuously, linked to documented review and corrective action paths, and communicated transparently to patients.[5,9,15,17]

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest: The authors declare no competing interests.

Acknowledgments: The authors have no acknowledgments to declare.

Data Availability: No new patient-level dataset was generated or analyzed. The literature review and practice-based synthesis use aggregate results reported in the cited publications.[11,16,17]

References:

1. Sakkas D. The “golden fleece of embryology” eludes us once again: a recent RCT using artificial intelligence reveals again that blastocyst morphology remains the standard to beat. Hum Reprod. 2025;40(1):4-8. doi:10. 1093/humrep/deae263.
2. Lee T, Natalwala J, Chapple V, Liu Y. A brief history of artificial intelligence embryo selection: from black-box to glass-box. Hum Reprod. 2024;39(2):285-292. doi:10. 1093/humrep/dead254.
3. Zabari N, Kan-Tor Y, Or Y, et al. Delineating the heterogeneity of embryo preimplantation development using automated and accurate morphokinetic annotation. J Assist Reprod Genet. 2023;40:1391-1406. doi:10. 1007/s10815-023-02806-y.
4. Berntsen J, Rimestad J, Lassen JT, Tran D, Kragh MF. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. PLoS One. 2022;17(2):e0262661. doi:10. 1371/journal. pone. 0262661.

5. Cohen J, Silvestri G, Paredes O, et al. Artificial intelligence in assisted reproductive technology: separating the dream from reality. Reprod Biomed Online. 2025;50(4):104855. doi:10. 1016/j. rbmo. 2025. 104855.
6. Enatsu N, Miyatsuka I, An LM, et al. A novel system based on artificial intelligence for predicting blastocyst viability and visualizing the explanation. Reprod Med Biol. 2022;21:e12443. doi:10. 1002/rmb2. 12443.
7. Palmer GA, Chavez-Badiola A, Valencia-Murillo R, et al. Can artificial intelligence-guided feedback improve embryologists" selection of euploid embryos based on morphology alone? Reprod Biomed Online. 2025;51(4):104990. doi:10. 1016/j. rbmo. 2025. 104990.
8. Yuan Z, Yuan M, Song X, Huang X, Yan W. Development of an artificial intelligence-based model for predicting the euploidy of blastocysts in PGT-A treatments. Sci Rep. 2023;13:2322. doi:10. 1038/s41598-023-29319-z.
9. Afnan MAM, Liu Y, Conitzer V, et al. Interpretable, not black-box, artificial intelligence should be used for embryo selection. Hum Reprod Open. 2021;2021(4):hoab040. doi:10. 1093/hropen/hoab040.
10. Sfontouris I, Nikiforaki D, Liarmakopoulou S, et al. P-280 Potential for improvement and current limitations of artificial intelligence (AI) for embryo selection: analysis of external validation data. Hum Reprod. 2022;37(suppl 1):deac107. 269. doi:10. 1093/humrep/deac107. 269.
11. Sergeev S, Diakova I, Nadirashvili L. Comprehensive AI-based system for KPI assessment in IVF laboratories: development and implementation. Medical Times. 2024;2(1):66-76. doi:10. 71419/mtggrc. 2024. 19.
12. Illingworth PJ, Venetis C, Gardner DK, et al. Deep learning versus manual morphology-based embryo selection in IVF: a randomized, double-blind noninferiority trial. Nat Med. 2024;30(11):3114-3120. doi:10. 1038/s41591-024-03166-5.
13. Salih M, Austin C, Warty RR, et al. Embryo selection through artificial intelligence versus embryologists: a systematic review. Hum Reprod Open. 2023;2023(3):hoad031. doi:10. 1093/hropen/hoad031.
14. Popa T, He C, Vasconcelos F, et al. P-168 Both artificial intelligence and manual embryo selection methods show sex-bias, favouring male embryos – insights from the largest embryo sex study using time-lapse and PGT-A. Hum Reprod. 2024;39(suppl 1):deae108. 539. doi:10. 1093/humrep/deae108. 539.
15. Koplin JJ, Johnston M, Webb ANS, Whittaker A, Mills C. Ethics of artificial intelligence in embryo assessment: mapping the terrain. Hum Reprod. 2025;40(2):179-185. doi:10. 1093/humrep/deae264.
16. Sergeev S, Diakova I. Advanced KPI framework for IVF pregnancy prediction models in IVF protocols. Sci Rep. 2024;14:29477. doi:10. 1038/s41598-024-80759-7.
17. Sergeev S, Diakova I, Nadirashvili L. Neural networks pipeline for quality management in IVF laboratory. J IVF-Worldwide. 2024;2(4):1-9. doi:10. 46989/001c. 124947.