

Original Research Articles

Advancing Bovine Embryo Evaluation: Machine Learning to Assess Embryos in Routine Embryo Transfer Practice

Cara Wells¹^a, Cameron Hayden¹, Michael Rea^{1,2}, Ben Cook^{1,3}, Russell Killingsworth^{1,4}

¹ EmGenisys, Inc., Driftwood, Texas, United States, ² Yellowstone Genetics, Billings, Montana, United States, ³ Sparrow Computing, Omaha, Nebraska, United States, ⁴ Shamrock Veterinary, Shamrock, Texas, United States

Keywords: bovine embryo transfer, embryo evaluation, artificial intelligence, in vitro fertilization

<https://doi.org/10.46989/001c.141131>

Journal of IVF-Worldwide

Vol. 3, Issue 3, 2025

Introduction

Machine learning (ML) enables rapid, high-precision data processing and holds promise for outperforming human evaluators in livestock embryo transfer (ET), ultimately accelerating genetic progress and improving economic outcomes. Therefore, this study had two primary objectives: (1) to evaluate ML performance in a real-world field trial, and (2) to survey embryologists on their traditional assessments of bovine embryos and compare their evaluations to ML-generated results.

Materials and Methods

1. A dataset of 6,900 30s videos of bovine embryos were recorded during routine ET with a smartphone mounted to a microscope and morphology was evaluated.
2. 42 bovine embryologists were surveyed to evaluate ten embryo images. Responses were compared to ML predictions.
3. 573 embryos compared ML stage and grade predictions compared to embryologists' results.
4. Kruskal–Wallis tests with Bonferroni corrections were used to assess differences in embryo assessments across groups, and independent t-tests were applied where assumptions of normality and equal variance were met.

Results

Embryologist assessments showed only 59.8% agreement overall, increasing to 74.6% among those with over 5 years of experience. ML demonstrated 70% agreement with all participants and 85% with experts, showing no statistical difference compared to expert evaluations ($p > 0.05$). ML was also proficient in identifying unfertilized oocytes, typically a skill of experienced embryologists. In the broader study, ML reached 81.7% agreement with experts on embryo stage (456/558) and 95.2% on transferability (531/558).

Conclusion

This study highlights new ML applications to evaluate embryos, offering the reality of an automated and standardized embryo analysis with the potential to improve status quo.

^a Corresponding author: Cara Wells, Email: cara@emgenisys.com

1. INTRODUCTION

1.1. MORPHOLOGICAL ANALYSIS AS THE GOLD STANDARD OF EMBRYO EVALUATION

Assisted reproductive techniques, including *in vitro* fertilization (IVF) and embryo transfer, (ET) in livestock species has yielded transformational genetic progress as these methods allow animal breeders, specifically beef and dairy producers, to maximize the genetics of both male and female animals. While the first successful pregnancy via embryo transfer was achieved for the first time in 1891 in rabbits, these technologies did not become available for use in routine practice in cattle until the 1970s.¹ These methods were initially offered by pioneering veterinarians, first as a surgical procedure and then later as a non-surgical procedure in 1976.²⁻⁴ This spurred the emergence of more commercial ET operations and IVF laboratories to make ET accessible to beef and dairy producers.

Nearly 50 years have passed since the advent of non-surgical ET and the International Embryo Technology Society (IETS) reports that over 1.1 million transferrable livestock embryos were produced in 2022, although actual numbers are much higher as reporting is voluntary.⁵ Scientific societies, such as the IETS and others, have held annual meetings to share progress and advancements in the ET industry, and have attracted membership from veterinarians, researchers, academics, and field technicians worldwide. These members have collectively accomplished many amazing scientific feats, including cloning and transgenics, but shockingly, few changes to embryo evaluation and selection have resulted for use in the commercial ET industry. Nearly 100% of the commercial ET industry relies on a simple evaluation of embryos under a microscope in which embryos are classified based on a number code system for their stage of development (1 to 9) and for their quality (1 to 4) (Table 1, Table 2).⁶ This morphological analysis, introduced in 1998, remains the gold standard for bovine embryo evaluation and grading within the bovine embryo production and transfer industry despite the widespread understanding that this morphological assessment is biased by the subjectivity of the evaluator and not considered 100% reliable or trustworthy.⁷⁻⁹

1.2. AI ROLE IN THE TECHNOLOGY REVOLUTION

In the nearly 50 years spanning the introduction of non-surgical ET and the increased use of *in vitro* embryo production to breed cattle, the technology revolution was gaining traction in parallel, specifically regarding artificial intelligence (AI). AI refers to the field of computer science focused on creating systems capable of performing tasks that typically require human intelligence such as learning, reasoning, problem-solving, perception, language understanding and decision-making.¹⁰ The term AI was coined in 1956, and later expanded to encompass machine learning (ML), computer vision, deep learning and convolutional neural networks which have the capability to evaluate numerical, language, and image data alike.¹¹⁻¹⁶

Advancements in computer processing and cloud hosting services further propelled the AI revolution and by the 2020s no industry has been left untouched by application of AI technologies, including healthcare, agriculture and even embryo evaluation. A recent PubGrade search of scientific literature using the keywords “embryo evaluation using artificial intelligence” shows an increasing number of publications each year with over 70 papers published in year 2024 alone. Special interest groups within reproductive medicine and fertility organizations have emerged to focus exclusively on the use of AI in reproductive medicine and meet regularly to disseminate research and drive policy. Use of AI in the human IVF laboratory has outpaced adoption of AI in the livestock IVF laboratory, as the field of human embryology now utilizes many AI enabled technologies to improve embryo selection, automate laboratory procedures, and enhance personalized patient care.

In a 2023 review of 20 studies evaluating AI to evaluate images of embryos, all studies reported that the AI outperformed the embryologist's evaluation of embryos in terms of embryo morphology assessments or reproductive outcomes, from 4-45%.^{17,18} These findings greatly support the use of AI to in the IVF laboratory as such advancements can increase live birth outcomes, reduce time to pregnancy and lower the financial burden of IVF.^{19,20} In 2024, Alife health completed the first US randomized control trial, which included 440 patients and showed improved ongoing pregnancy outcomes when using AI-enhanced embryo selection compared to traditional morphology grading alone.²¹ Other studies report similar findings, such as work performed by Wang *et al* in which AI-assisted embryo selection led to a higher implantation rate (80.87%) compared to manual selection (65.15%), without compromising neonatal outcomes.²² Other studies have published promising results showcasing the use of AI for euploid prediction to provide a potential non-invasive, efficient, and cost-effective tool for embryo selection with some suggesting that non-invasive AI analysis can improve live birth outcomes, even for embryos classified as euploid by traditional PGT-A methods.²³⁻²⁷

Based on the success of AI to add value to the human IVF industry, integrating AI into livestock ET practice could allow advanced, automated, embryo analysis to accurately predict embryo viability, and surpass the ability of human evaluators to break barriers currently restricting the potential of ET to generate genetic advancement and limiting economic returns.²⁸

1.3. EARLY APPLICATIONS OF AI INTO LIVESTOCK ET

While AI certainly holds the potential to break barriers and improve the status quo of livestock ET, initial work utilizing AI to evaluate livestock embryos compared to the current gold standard of embryo evaluation, the morphological analysis, is warranted. To explore these capabilities, the aim of this study was two-fold: 1) Train machine learning models to predict embryo stage and quality grade based on 30s video data captured with standard microscopy and imaging equipment in a field trial and 2) Survey embryolo-










Stage Code	Description	Image
1	Unfertilized oocyte	
2	2- to 12-cell	
3	Early morula	
4	Morula	
5	Early blastocyst	
6	Blastocyst	
7	Expanded blastocyst	
8	Hatched blastocyst	
9	Expanded hatched blastocyst	

Table 1. Stage of Development. Description of codes associated with embryo stage of development with associated images.

gists’ assessment of bovine embryos with traditional methods and compare ML results.

2. MATERIALS AND METHODS

2.1. TRAINING THE MODEL

Original methods to detect and evaluate bovine embryos from video data was described in Wells *et al.*¹⁸ To train new models to evaluate embryo developmental stage and quality grade, 6,900 30s videos of both *in vivo* derived and *in vitro* produced bovine embryos were recorded from ten ET practitioners during routine ET. Videos were recorded with equipment that the ET practitioners owned prior to the

study and included several microscope models and camera types. Each video was recorded at a magnification of 90x using a stereoscope coupled with a 3x optical zoom on the camera, resulting in an overall magnification of 270x ([Figure 1](#)). A maximum of 12 embryos were recorded in each video, with careful orientation to maintain a minimum distance of 10 µm between embryos, ensuring that they did not touch or overlap ([Figure 1](#)). Precautions were taken to prevent embryo drift and external noise, including the use of a level laboratory table, and measures were implemented to minimize external disturbances, such as silencing fans, air conditioning, and radio equipment to avoid interference from external stimulus.

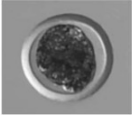
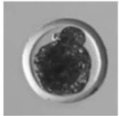

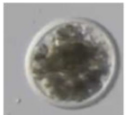
Quality Grade Code	Description	Image	Traits and Characteristics
1	Excellent or Good		The embryo is well-defined, spherical in shape with no visible defects
2	Fair		The embryo shows minor irregularities in shape or cell division, extruded cells, or fragmentation
3	Poor		The embryo shows irregularities and fragmentation that consume greater than 50% of the embryos volume
4	Dead or Degenerating		The embryo is poorly developed with major fragmentation, uneven cell size and visible defects

Table 2. Quality Grade. Description of codes associated with embryo quality grade.

Embryo images depict stage 4 morulas of varying quality grades.

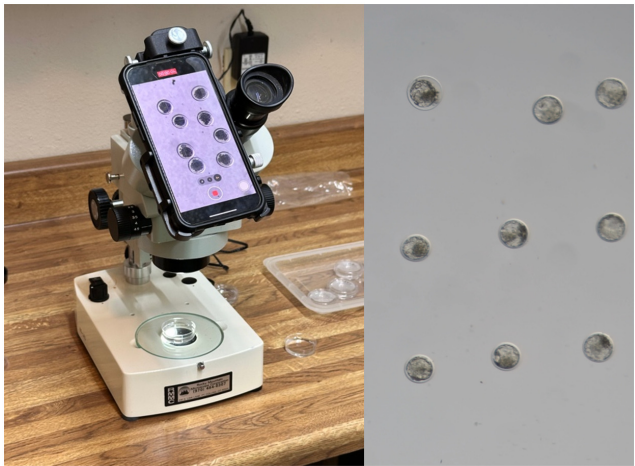


Figure 1. Image of phone mounted to microscope used to record videos and screen capture of bovine embryos recorded at 270x and oriented 10um apart.

This represents equipment used to capture videos used by bovine embryo transfer practitioners.

Because the data collection occurred during routine embryo transfer and involved no experimental manipulation of live animals, no Institutional Animal Care and Use Com-

mittee (IACUC) approval was required. The video capture and analysis procedures were entirely non-invasive and did not involve altering or physical interaction with the embryo. As such, this study was exempt from IACUC oversight. Furthermore, because the ML model analyzes pre-existing video data and does not influence embryo handling, selection, or treatment, or alter any material entering the human food supply, it is not subject to regulatory oversight by the USDA, FDA, or other regulatory body under current guidelines for bovine ET practices.

During the time of video capture, each embryo was evaluated according to IETS standards to include a development stage and quality grade. Development stage and quality grade will constitute as ML labels for ML training. Embryo evaluation methods are described in detail by the IETS and aim to classify embryo development and quality with a series of codes. Pre-implantation embryo stage of development is represented by codes 1-9, in which 1 represents an unfertilized oocyte, the lowest stage of development and 9 represents an expanded hatched blastocyst, which is an embryo that consists of differentiated cell types that has undergone expansion, hatched from the zona pellucida, and it preparing to implant into the endometrium (Table 1). Embryo quality grade is represented by codes 1-4, in which 1 represents an “excellent or good embryo” that

is well-defined, spherical in shape with no visible defects. This is considered the highest quality and most likely to lead to successful pregnancy. Grade 4 is a dead or degenerating embryos that is poorly developed with major fragmentation, uneven cell size and visible defects which are known to have a low chance of successful implantation or pregnancy and are generally not recommended for transfer (Table 2).

Videos were then uploaded in CVAT (Sparrow Computing, Nebraska, USA) to train the model for object detection, which refers to the task of identifying and locating objects withing an image or video. Within CVAT, bounding boxes were drawn around each embryo. This task aimed to train ML models to detect and recognize embryos apart from other objects and debris. Next the model was validated and tested to determine proficiency at detecting and recognizing embryos. Next videos were imported into EmGenisys EmVision Software (Driftwood, TX, USA) in which the developed object detection model was hosted. In EmGenisys EmVision Software, detected embryos were labeled with a development stage and quality grade label. All data was stored in EmGenisys EmVision Software database, hosted by Amazon Web Services (AWS). Here, the labeled data was used to train the ML models, using a selection of image-based deep learning classification tools. Models were then validated to prevent over and underfitting and later tested on real-world samples of videos of embryo which were not included in the original training or validation sets.

2.2. SURVEY OF EMBRYOLOGISTS

Ten images of embryos, which were derived from screenshots of video images of embryos with known labels, were selected for use in a survey. Images were selected to encompass ten embryos of varying stages and grades. Images were uploaded into Jotform, an online survey platform. The survey was ongoing in January-February 2024, spanning the IETS annual meeting in Denver, CO. During this conference, attendees were asked to take the survey.

The survey included both demographic information about the respondent, including years of experience, role in bovine embryology, and location. Then, respondents were asked to evaluate the ten images of embryos and classify them according to stage and grade. Once the survey was closed, the 30s videos of the ten embryos images included in the survey were processed with the ML platform to create a computer-generated stage and grade. Then, the stage and grade of the survey respondents' classifications were compared to the computer-generated classifications.

To evaluate the impact of embryologists' experience level on embryo scoring from the survey results, a total of 42 embryologists were recruited to assess the ten embryos. Embryo stage was scored on a scale of 1 to 9 based on IETS standards. Residuals were calculated by subtracting the observed embryo score from the corresponding embryologist-assigned score. These residuals were then assessed for normality using the Shapiro-Wilk test. The Shapiro-Wilk test revealed a significant deviation from normality ($W = 0.7605$, $p < 0.001$), indicating that the residuals were not normally distributed. A visual inspection of the residuals using a his-

togram and Q-Q plot further supported the non-normal distribution, characterized by a pronounced concentration of zero values, representing exact matches between observed and assigned scores.

Given the violation of the normality assumption, a Kruskal-Wallis test was employed to compare the embryo scores across experience levels. The Kruskal-Wallis test was selected as a non-parametric alternative to ANOVA, as it does not assume normally distributed data. Data were grouped by experience level (e.g., novice, experienced, expert), and tested per embryo to determine if differences in embryo scoring were statistically significant. Post-hoc analysis was conducted using Dunn's test with Bonferroni correction to identify specific group differences, with significance defined as $p < 0.05$ and an effect size of 0.3. This test was performed individually for each of the ten embryos, resulting in a total of ten separate analyses.

Similar methods were applied to assess differences in embryo quality grade, a qualitative score ranging from 1 to 4 based on the IETS standards.

2.3. EXPANDING ACROSS A LARGER DATASET

To better assess accuracy of the ML model, three embryologists with at least ten years of experience evaluating embryos scored a total of 558 bovine embryos. Embryos were only scored by one embryologist each, which is recognized as a limitation of this study. Embryos were then passed through the ML model to create a computer-generated embryo stage and grade code. Computer generated stage and grade codes were compared to the scores of the embryologists' labels. Accuracy was determined based on agreement between computer generated stage and grade code as both an exact match to the embryologists' labels as well as agreement between computer generated stage and grade codes ± 1 from the embryologist labels.

To evaluate the agreement between the machine learning (ML) model and the expert embryologist in classifying embryo grades, Cohen's Kappa (κ) statistic was used. This statistic measures inter-rater reliability while accounting for agreement that could occur by chance. Embryo grades were assigned using a discrete, ordinal scale, and two forms of Kappa were calculated to assess model performance.

The unweighted Cohen's Kappa was first calculated to measure exact agreement, treating all disagreements equally regardless of how far apart the categories were. Because the grading system is ordinal, a weighted Cohen's Kappa was also computed using linear weights to account for the relative distance between mismatched grades. For example, a disagreement between stage 6 and 7 is penalized less than a disagreement between stage 1 and 6. The number of observed agreements and the number expected by chance were used to calculate the kappa values, along with the standard error and 95% confidence intervals.

Interpretation of κ followed the classification scale proposed by Landis and Koch, which defines ranges of agreement from slight to almost perfect. All statistical calculations were performed using GraphPad QuickCalcs, an online tool provided by GraphPad Software. The tool supports both unweighted and linear weighted Kappa com-

putations and was used to ensure consistent and accurate assessment of inter-rater agreement. Interpretation of κ followed the scale proposed by Landis and Koch, which categorizes agreement as slight, fair, moderate, substantial, or almost perfect based on the value of κ .

3. RESULTS

3.1. SURVEY RESPONDENT DEMOGRAPHICS

A total of 42 ($n=42$) embryologists completed the survey. Embryologists' experience ranged from 1 to 40 years: 13 (31) had 0-5 years, 16 (38%) had 5-10 years, and 13 (31) had greater than 10 years' experience working in bovine embryology. Survey respondents were also asked information regarding their role in bovine embryology. 9/42 (21%) reported working in academics, 14/42 (34%) reported working as an embryologist, and 19/42 reported working as a veterinarian. Respondents that reported working in academics held either a master's or PhD degree in reproductive biology, embryologists worked in industry without a veterinary degree, and veterinarians included those working exclusively in ET and non-exclusively in ET as a bovine veterinarian. Survey respondents worked in a total of six countries: the United States, Brazil, India, Turkey, Colombia and Australia.

3.2. SURVEY RESPONDENT AND MACHINE LEARNING EMBRYO ASSESSMENT

The survey revealed significant disparities in embryo stage assessments among embryologists of different experience levels ($p<0.05$), with only 59.8% agreement across all participants. Agreement classifying embryo stage notably increased to 74.6% among "Experienced" (5-10 years) and "Expert" (>10 years) embryologists. Agreement classifying embryo stage amongst all embryologists was greater than 50% for 6/10 embryos. Only 1/10 embryos demonstrated >75% agreement amongst all embryologists, which included the stage 8 hatching blastocyst (Table 3).

In contrast, ML demonstrated 70% agreement with all participants and 85% agreement specifically with "Expert" embryologists, showing no statistical difference compared to expert embryologists ($p>0.05$). Notably, ML matched or exceeded embryologists' proficiency in identifying the unfertilized oocyte (Table 3).

Kruskal-Wallis Test was performed on agreement per embryo, based on embryologist's experience level and the ML model. 4/10 embryos demonstrated statistically significant differences in the embryologist's assessment of the embryo stage ($p<0.05$). Embryos which demonstrated statistical differences across study groups included a stage 7 blastocyst, a stage 8 hatching blastocyst, a stage 9 hatched blastocyst, and a stage 1 unfertilized oocyte. No statistical differences were found between the ML model and expert embryologists with more than ten years of experience for any of the ten embryos (Table 3).

When assessing embryo grade, statistical differences were found between study groups for 5/10 embryos. Em-

bryos which demonstrated a statistical difference included a grade 4 early morula (stage 3), a grade 1 hatching blastocyst (stage 8), a grade 1 hatched blastocyst (stage 9), a grade 4 unfertilized oocyte (stage 1), and a grade 1 morula (stage 4) (Table 3).

3.3. MACHINE LEARNING ASSESSMENT OF BOVINE EMBRYOS FROM A LARGER DATASET

In the broader study, the ML model achieved (456/558) 81.7% agreement with the expert embryologist identifying embryo stage. In the event the ML model and embryologist did not agree on embryo stage, (132/141) 93.6% of predictions were 1 stage apart, typically in disagreement over stage 6 and stage 7 embryos showing that the ML model again, evaluates embryos comparatively to expert embryologists ($p=0.5$) (Table 4).

To assess inter-rater agreement, Cohen's Kappa (κ) was calculated based on exact matches between the expert embryologist assigning the labels the ML prediction. The number of observed agreements was 456 out of 558 observations, corresponding to 81.72% agreement. The number of agreements expected by chance was 171.1 (30.66%). The resulting unweighted Cohen's Kappa was $\kappa = 0.736$, with a standard error (SE) of 0.023 and a 95% confidence interval ranging from 0.691 to 0.782, which is considered to be substantial agreement. Because the classification categories were ordinal, weighted Cohen's Kappa was also calculated using linear weights, which takes into account the relative distance between mismatched categories. The resulting weighted Kappa was $\kappa = 0.855$, indicating almost perfect agreement between the ML and the embryologist. These findings suggest a high level of consistency between raters, particularly when accounting for the ordered nature of the classification system (Table 4).

To evaluate the agreement between the machine learning (ML) model and the expert embryologist for embryo grade classification, Cohen's Kappa (κ) was used. Exact matches between the ML-predicted grades and the embryologist's assessments were observed in 416 out of 557 cases, corresponding to 74.69% agreement (Table 5). The number of agreements expected by chance alone was 246.6 (44.28%). The resulting unweighted Cohen's Kappa was $\kappa = 0.546$, with a standard error (SE) of 0.033 and a 95% confidence interval from 0.481 to 0.610, indicating moderate agreement between the ML model and the expert. Because embryo grades are ordinal in nature (e.g., 1,2,3,4.), where adjacent grades are more similar than distant ones, a weighted Cohen's Kappa using linear weights was also calculated. This approach accounts for the degree of disagreement based on how far apart the predicted and actual grades are. The weighted Kappa was $\kappa = 0.684$, reflecting a substantial level of agreement between the ML model and the embryologist when the ordinal structure of the grading system is considered (Table 5).

While not to discount the credibility of the statistical analysis, it should be noted that the main disparity in agreement is between embryo quality grade 1 and 2. As most ET practices will transfer both quality grade 1 and 2 embryos alike, as it is well known that both quality grades











Years Experience <5 (Novice) 5-10 (Experienced) >10 (Very Experienced) ML							Kruskal-Wallis Test
	Embryo 1	Stage	5.8 ^a	6.9 ^b	6.9 ^b	7 ^b	Kruskal-Wallis Test (stage): p<0.05; H=8.4; Effect size η^2 =0.16 (large)
	Embryo 1	Grade	1.9	1.5	1.4	A (1-2)	Kruskal-Wallis Test (grade): p>0.05; H=3.2; Effect size η^2 =0.032 (small)
	Embryo 1	Mode	7--1	7--1	7--1		
	Embryo 2	Stage	3.9	3.1	3.1	4	Kruskal-Wallis Test (stage): p>0.05; H=3.2; Effect size η^2 =0.03 (small)
	Embryo 2	Grade	2.3 ^a	3.2 ^a	3.5 ^b	C (4)	Kruskal-Wallis Test (grade): p<0.05; H=12.6; Effect size η^2 =0.28 (large)
	Embryo 2	Mode	3--3	3--4	3--4		
	Embryo 3	Stage	7.5 ^a	8 ^b	8 ^b	8 ^b	Kruskal-Wallis Test (stage): p<0.05; H=6.7; Effect size η^2 =0.12 (medium)
	Embryo 3	Grade	1.7 ^a	1.0 ^b	1.3 ^b	A (1-2)	Kruskal-Wallis Test (grade): p<0.05; H=5.1; Effect size η^2 =0.081 (medium)
	Embryo 3	Mode	8--1	8--1	8--1		
	Embryo 4	Stage	5.3	6.7	6	7	Kruskal-Wallis Test (stage): p>0.05; H=3.7; Effect size η^2 =0.045 (small)
	Embryo 4	Grade	2.7	2.1	2.4	A (1-2)	Kruskal-Wallis Test (grade): p>0.05; H=1.23; Effect size η^2 =-0.02 (very small)
	Embryo 4	Mode	7--3	7--2	7--2		
	Embryo 5	Stage	6.4 ^a	8.2 ^b	8.6 ^b	7 ^{a,b}	Kruskal-Wallis Test (stage): p<0.05; H=6.2; Effect size η^2 =0.11 (medium)
	Embryo 5	Grade	1.6 ^a	1.3 ^b	1.2 ^b	A (1-2)	Kruskal-Wallis Test (grade): p<0.05; H=9.3; Effect size η^2 =0.19 (large)
	Embryo 5	Mode	9--2	9--1	9--1		
	Embryo 6	Stage	4.3 ^a	3 ^a	1.5 ^b	1 ^b	Kruskal-Wallis Test (stage): p<0.05; H=12.104; Effect size η^2 =0.26 (large)
	Embryo 6	Grade	1.9 ^a	2.3 ^a	3.5 ^b	C (4)	Kruskal-Wallis Test (grade): p<0.05; H=6.83; Effect size η^2 =0.13 (medium)
	Embryo 6	Mode	4--1	4--1	1--4		
	Embryo 7	Stage	4.9	5.9	5.8	7	Kruskal-Wallis Test (stage): p<0.05; H=3.5; Effect size η^2 =0.39 (small)
	Embryo 7	Grade	2.3	1.4	1.6	A (1-2)	Kruskal-Wallis Test (grade): p>0.05; H=4.6; Effect size η^2 =-0.07 (medium)
	Embryo 7	Mode	5--2	6--1	6--1		
	Embryo 8	Stage	4.6	4.3	4.2	4	Kruskal-Wallis Test (stage): p>0.05; H=3.5; Effect size η^2 =0.039 (small)
	Embryo 8	Grade	1.9 ^a	1.4 ^{a,b}	1.2 ^b	A (1-2)	Kruskal-Wallis Test (grade): p<0.05; H=8.3; Effect size η^2 =0.17 (large)
	Embryo 8	Mode	4--2	4--1	4--1		
	Embryo 9	Stage	5.6	4.9	4.9	5	Kruskal-Wallis Test (stage): p>0.05; H=3.0; Effect size η^2 =0.026 (small)
	Embryo 9	Grade	2.5	1.8	2.2	A (1-2)	Kruskal-Wallis Test (grade): p>0.05; H=0.4; Effect size η^2 =-0.04 (very small)
	Embryo 9	Mode	5--1	5--2	5--2		
	Embryo 10	Stage	4.6	4.2	4.1	4	Kruskal-Wallis Test (stage): p>0.05; H=71.49; Effect size η^2 =-0.013 (very small)
	Embryo 10	Grade	2.5	2.7	2.8	A (1-2)	Kruskal-Wallis Test (grade): p>0.05; H=3.8; Effect size η^2 =-0.05 (small)
	Embryo 10	Mode	4--2	4--3	4--3		
Stage priori power is low (0.23), Grade priori power is low (0.25)							

Table 3. Survey outcomes of embryologists' classification of bovine embryos based on years of experience compared to machine learning prediction.

A-b denotes significance. Color transitions were used to visually represent the degree of consensus or discrepancy in embryo evaluation across different embryologist experience levels. Green represents the highest scores for each assessment category, red represents the lowest score for that category, and yellow represents intermediate scores between highest and lowest. The shading of each color reflects the effect size and assessment variability. For example, green to light green shows a small effect size where embryologists gave similar scores, green to yellow to red to yellow shows a medium effect size where there is moderate disagreement between embryologists of different experience levels and green to red shows a high effect size with substantial disagreement.

Table 4. Agreement between Machine Learning prediction and actual stage labels.

Stage	1	2	3	4	5	6	7	8	9
Exact Match	18/21 (85.71%)	17/22 (77.27%)	0/0 -	30/36 (83.33%)	17/29 (58.62%)	163/196 (83.16%)	198/234 (84.62%)	9/15 (60%)	4/5 (80.0%)
+/- 1	20/21 (95.24%)	21/22 (95.45%)	0/0 -	34/36 (94.44%)	28/29 (96.55%)	196/196 (100%)	233/234 (99.57%)	13/15 (86.67%)	5/5 (100.0%)

Color transitions were used to visually represent the degree of consensus or discrepancy in embryo evaluation across the expert embryologist and Machine Learning prediction. Agreement >90% - dark green, 80-89% - green, 70-79% yellow-green, 60-69% - yellow, 50-59% - orange, <50% - red.

can consistently produce pregnancies and that discrepancy between embryo evaluators is prevalent, results were then organized to evaluate ML proficiency at predicting embryo transferability. When quality grades were grouped based on transferability, marked as embryo quality grade 1-2= Transfer, 3=Marginal and 4=Non-transferrable, the ML model achieved (531/557) 95.3% agreement (Table 5).

4. DISCUSSION

4.1. SURVEY OUTCOMES: EMBRYO EVALUATION DISCREPANCIES AMONG EMBRYOLOGISTS

It is well-established that embryo evaluation is one of the most critical factors contributing to the success of embryo transfer, yet the industry standard relies on antiquated, subjective morphological methods which are known to have poor accuracy, poor reproducibility and to be biased by the

Table 5. Agreement between Machine Learning prediction and actual grade labels.

Grade	1	2	3	4
Exact Match	269/351 (76.64%)	96/149 (64.43%)	2/4 (50.0%)	49/53 (92.45%)
+/- 1	344/351 (98.01%)	148/149 (99.33%)	3 / 4 (75.0%)	52/53 (98.11%)

Color transitions were used to visually represent the degree of consensus or discrepancy in embryo evaluation across the expert embryologist and Machine Learning prediction. Agreement >90% - dark green, 80-89% - green, 70-79% yellow-green, 60-69% - yellow, 50-59% - orange, <50% - red.

subjectivity of the evaluator.^{7,8,29-31} These fundamental, industry fallacies were further emphasized by survey results, as only 59% of embryologists agreed on embryo stage and significant differences amongst evaluators existed on 5/10 embryos when evaluating embryo quality grade. Real-world consequences of these discrepancies vary from minor to catastrophic, sometimes yielding major economic burdens and decreased consumer trust in the ET industry.

For minor discrepancies, which is most cases, the discrepancy is simply identifying embryo stage or grade +/- 1 from actual result. For example, a stage 6 embryo is defined as a blastocyst and a stage 7 embryo is defined as an expanded blastocyst. Expansion is usually evidenced by increased embryo diameter and a thin, stretched zona pellucida. In the survey, Embryo-4 represented a blastocyst in which the zona was stretched thin and of increased diameter, but the embryo proper was collapsed did not consume the full subzonal space. Average stage prediction for this embryo was 6.7 for experienced embryologists, and 6.0 for expert embryologist, though the mode for both experienced and expert embryologists was stage 7 (Table 3). This data underscores how even experts are split in uniformly identifying embryos and struggle to accomplish uniformity when embryo morphology is not a textbook example of normal embryo development and progression. However, in this case example, the economic impact of misidentifying the embryo stage is nominal. For bovine IVF embryos, most embryologists consider both 6 and 7 embryos acceptable for transfer and would not discard the embryo either way. While the intent of the embryo morphological evaluation system is to standardize the process, the commercialization of the embryo evaluation system can inject bias into this standard. Commercially, some IVF companies capture a premium for stage 7, grade 1 embryos as embryos of this morphological classification are known to product the highest pregnancy outcomes.

In this same case example, average stage prediction for Embryo-4 by novice embryologists was 5.3 (early blastocyst) (Table 3). It is well known that embryos that have not progressed into well-formed blastocysts by day 7 post-fertilization have likely stalled and will not continue to develop. In these cases, many embryologists would elect to discard these embryos as they cannot be expected to ever result in full-term pregnancy. With this logic, this embryo would have been discarded by 11/42 of the embryologists, which can have significant economic impacts such as missed opportunity of producing a valuable calf.

Conversely, Embryo-6, presents an interesting case study. Admittedly, Embryo-6 is not in fact an embryo as it is a stage 1 unfertilized oocyte. Characterized by a single cell

with smooth edges encased in an intact zona pellucida, this unfertilized oocyte can appear to be similar to a compacted morula (stage 4) to the untrained eye. The average and mode assessment for this cell by Novice embryologists was average stage 4.3 (morula), average grade 1.9 (high quality and transferrable) and mode 4-1 (stage 4, grade 1). Even most experienced embryologists failed to recognize this unfertilized oocyte, as the mode for experienced embryologist was also 4-1. Only expert embryologists accurately identified this Stage 1 unfertilized oocyte consistently (Table 3).

As this example was produced during *in vivo* embryo collection, a stage 4 morula is typically considered acceptable for freeze and transfer. 25/42 of the respondents of this study believed this cell to be a morula and would have included it in the ET program. The transfer of this unfertilized oocyte has a 0% chance of resulting in a pregnancy or live calf, and this lack of oversight would have placed an undue burden on the client to cover the cost of embryo freezing, transfer, and recipient care. As the incidence of collecting unfertilized oocytes is extremely common during embryo transfer, the inability to accurately identify these cells can result in a compounding economic problem over time which drastically reduces the return on investment of ET to the beef and dairy industry.

4.2. SURVEY OUTCOMES: EMBRYO EVALUATION AGREEMENT INCREASES WITH EXPERIENCE

Survey results show that agreement significantly increases with experience, as embryologists with >10 years of experience demonstrate more uniformity when evaluating embryos. These results are expected as most skills are perfected over time and proficiency typically increases with experience. No significant differences in identifying embryo stage were present between experienced embryologists (5-10 years) and expert embryologists (>10 years) for any of the 10 embryos included in the survey. It is the Novice embryologists, with less than 5 year's experience, who demonstrated significant differences in evaluating both embryo stage and grade at a higher frequency than their experienced and expert counterparts ($p < 0.05$). While it is widely acknowledged that novice embryologists are learning and cannot be penalized for making mistakes, this inefficiency stifles the ET industry and reduces economic return for cattle producers. Despite this, the potential of ET to advance genetic gains and enhance competitiveness for beef and dairy producers continues to drive the demand of ET as a method to breed cattle and creates increased demand for embryologists to do this work. Therefore, methods to decrease training time of Novice embryologists and quickly increase embryologist proficiency at evaluating embryos is

valuable.

It should be noted that a limitation of this study includes that survey respondents were recruited to participate during an academic conference. Many of the novice embryologists were graduate students and not yet practicing in the field; however, they are not exempt from industry relevance, as many will soon be seeking employment and transitioning into industry roles. On the other end of the spectrum, the demographic recruited also included individuals willing to invest thousands of dollars in registration and travel expenses to attend the conference—typically in pursuit of continuing education or to expand their knowledge in embryology. Many of the expert respondents were well-published professionals, recognized key opinion leaders, and actively engaged in the field, often with a strong interest in learning about new biotechnologies. For these reasons, the authors acknowledge that the demographics do not fully represent individuals performing bovine embryo transfers on farms. However, respondent demographics were reported (embryologist, veterinarian, academic), were evenly distributed, and these factors were taken into consideration in the interpretation of results.

4.3. THE ROLE OF MACHINE LEARNING TO EVALUATE EMBRYOS

The ML described by these researchers is based on the training of models to evaluate embryos stage and grade based on labeled embryo images provided in supervised datasets. While the underlying premise of ML is founded upon complex mathematical algorithms, methods to train these models in this study was straightforward. Supervised data in this context means that each data point or specifically, the video of each embryo, has a corresponding label or output (embryo stage and grade). This labeled data can then be used to train ML algorithms to identify patterns and relationships between input features (image data) and corresponding outputs (labels). Once trained, the algorithm can then be used to predict the output for new, unseen data, based on the learned patterns.³¹

Prior to this study, a ML model was trained on video data of 6,900 bovine embryos with stage and grade labels. These labels were provided by an expert embryologist with >10 years of experience in bovine embryology in attempt to most accurately train the models. However, because only one individual provided the training labels, it also means the model may have inherited unconscious bias or subjective interpretation inherent to a single evaluator. Initially, model performance is expected to reflect the nature of the dataset and perform comparably to the human embryologist assigning the labels. This means that the model will likely stage and grade embryos similarly to the expert who labeled the training set. However, as the model is improved with increased data and advancements in ML like generative artificial intelligence are applied, it can be expected that the model can improve and potentially outperform even expert human evaluators. At the time of the study, these generative artificial intelligence advancements were not included in the model. To address the limitations associated with subjectivity and improve model generaliz-

ability, future studies should incorporate labels from multiple independent embryologists to quantify inter-evaluator variability and reduce the influence of individual bias. Additionally, implementing ensemble labeling strategies or probabilistic modeling approaches could further enhance the reliability of the ground-truth data by capturing evaluator disagreement and modeling label uncertainty. This would allow the model to learn from a broader spectrum of expert opinion rather than relying on a single source of truth.

To test the model performance, it is critical tests be performed on new data that was not included in the training set of 6,900 embryos. The 10 embryos featured in the survey and 557 embryos featured in the expanded study were not included in the training set, and represent new, unseen data that the model had not been exposed to in effort to represent real-world scenarios that are likely to be encountered by bovine embryologists. Because of the subjective nature of embryo evaluation, it is inappropriate to present study results based on accuracy. Therefore, “agreement” and mode were selected to evaluate ML model performance.

While the subjective nature makes it difficult to proclaim that an embryologist evaluated the embryos in the survey correctly, it is expected that the expert evaluators can more proficiently apply the IETS evaluation standards than novice evaluators. The embryologists that labeled the embryos in the training set meet the criteria for expert embryologists in this study, as defined by as years of experience. When comparing expert embryologists survey responses and ML prediction of the 10 embryos included in the survey, no statistical differences were found between the ML model embryo evaluation and expert embryologists for either stage or grade, suggesting that ML can evaluate bovine embryos with comparable proficiency to that of an expert ($p>0.05$). Releasing this model into a web-hosting platform on local device could serve as a tool for embryologists to classify embryos to enhance the proficiency and boost the confidence of the novice embryologists, thus allowing novice embryologists to assess embryos comparably to someone with more experience.

In the expanded study, the ML model predicted the stage and grade of 558 embryos which had been labeled by one of three expert embryologists. While the embryologists evaluating the embryos each had acceptable qualifications to validate their skillset (education and years of experience), it is well established that even highly experienced embryologists do not always agree on both the stage and quality grade of the embryos due to the subjective nature of the morphological grading system.

This well-known observation was further supported in the survey where agreement amongst expert evaluators achieved 79% agreement, as well as in other cited literature. For example, Farin *et al* reports that when six experienced individuals evaluated 40 bovine embryos, agreement within evaluators for stage (89.2%) was higher than quality grade (68.5%) and that agreement among evaluators for stage was slightly higher for *in vivo* derived embryos (85%) than *in vitro* produced (72.3%).³⁰ With this knowledge, we can as-

sume there could be incongruity on the labeled data that went into the training model. Additionally, there is expected to be incongruity of the labels of the 558 embryos which were used to test the model. Therefore, results were reported both as an exact match and ± 1 stage code and ± 1 quality grade code to better represent real-world outcomes.

Agreement between the labeled data and ML prediction ranged from 58.62% to 85.71% for an exact match determining embryo stage, with mean agreement of 76.59% (Table 4). Agreement significantly increased when including agreement for 1 stage code difference, ranging from 86.67% to 100%, with mean agreement of 95.99% ($p < 0.05$) (Table 4). Agreement between the labeled data and ML prediction ranged from 50% to 92.45% for an exact match determining embryo quality grade, with mean agreement of 70.88% (Table 5). Agreement significantly increased when including agreement for 1 quality code difference, ranging from 75.0% to 99.33%, with a mean agreement of 92.61%. As the economic consequence of classifying embryos either 1 stage code or 1 quality code off is nominal, these outcomes support the use of utilizing this ML tool to support novice embryologists' evaluation of bovine embryos. Importantly, the ML demonstrated excellent proficiency at properly identifying non-transferrable embryos (stage 1-2 or quality grade 4), which can have an economic benefit to the ET industry (Table 5). Such tools break down barriers to entry to novice embryologists, who often require years of supervision before performing ET independently or fail to achieve pregnancy results comparable to more experienced embryologists, which have a negative impact on their reputation and limits their ability to secure clients. As the demand for ET in livestock outpaces embryologists, ML tools to improve embryologists' performance can make ET services more accessible to livestock producers and ultimately grow the ET industry.

4.4. USE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN THE LIVESTOCK EMBRYO TRANSFER INDUSTRY

The successful application of machine learning to stage and grade bovine embryos marks a pivotal step toward integrating AI into livestock embryo evaluation. Much like how AI/ML has been implemented into animal agricultural to detect pre-clinical signs of disease, detect mastitis from milk samples, and allow for animal identification through facial recognition to help farmers reduce costs, enhance performance, and play a decisive role in helping farmers see patterns and solutions to pressing problems in the modern animal agricultural industry, the application of AI/ML to evaluate embryos can play a role to improve efficiency and return on investment of ET.³² This foundational use case introduces standardization and consistency to bovine ET—an area that has seen limited innovation in the last fifty years—and opens the door to a broader transformation in how we assess and select embryos. While this study focused on classifying embryo stage and grade using accessible tools like stereomicroscopes and smartphones, it establishes a practical and scalable framework that can be

used in real-world, rural settings where technological infrastructure is often limited. With known pregnancy outcomes available for a portion of the embryos, future research will build on this work to train models for predicting embryo viability, pregnancy potential, early embryonic loss, and sex. Such advances could allow ET and IVF to rival the success rates of artificial insemination or live cover, enhancing both genetic progress and the economic return on assisted reproductive technologies.

However, for ML to be fully integrated into veterinary practice, it must remain accessible, cost-effective, and relevant to the realities of livestock producers working within narrow profit margins. Ethical considerations must remain at the forefront as this technology matures—particularly around genetic selection, biodiversity, and transparency in algorithm design. These tools should exist as a complement to the ET program and be used in a way that drives the industry and humans forward, rather than a tool that replaces technicians or promotes deskilling (a phenomenon that may lead to the deterioration of skills due to the reliance of AI/ML).³³⁻³⁵ As we expand into new species and more complex predictive models, it is critical that training datasets are diverse, free of technician bias, and subjected to peer-reviewed validation. Ultimately, the goal is not only to improve reproductive outcomes, but to do so in a way that supports animal health, producer equity, and environmental sustainability. Machine learning represents a powerful tool, but its responsible deployment will require collaboration between scientists, veterinarians, and producers to ensure its benefits are realized across the livestock industry.

5. CONCLUSION

This study further emphasizes challenges of embryo evaluation and discrepancies present among embryologists with varying experience levels; barriers that limits entry into the field and reduce producer access to ET as a tool for accelerating genetic gains. By demonstrating the use of ML to analyze 30-second real-time videos of bovine embryos and accurately classify developmental stage and grade, this study presents the first documented application of real-time video analysis for evaluating bovine embryos. Unlike static image-based approaches, video analysis allows for a more dynamic and nuanced assessment of embryo morphology and behavior, offering potential insights into embryo quality that may not be visible in still images.

Although this study successfully trained ML models using data collected during routine embryo transfers, we recognize its limitations—most notably, the relatively small group of researchers involved in labeling and training the models, and the need for a more diverse and representative dataset. Future work will actively seek collaboration with academic and research institutions to broaden the scope and rigor of these studies. We believe that ML will become a mainstay in the animal breeding industry, particularly within assisted reproductive technologies, much as it already has in human IVF programs. Next steps will extend this research to include embryo viability prediction and

long-term outcomes such as pregnancy and live birth rates—critical benchmarks needed to validate and refine these tools to meet the gold standard of ET success. Furthermore, implementing ML in ways that are affordable and accessible such as leveraging widely available tools like smartphones and stereomicroscopes will ensure that this technology can be applied across the livestock industry, even in rural and resource-limited settings. By reducing the subjectivity of embryo assessment, shortening the training curve for new embryologists, and improving consistency across practitioners, ML has the potential to democratize access to high-quality genetics and advance animal health, welfare, and productivity. Ethical deployment, transparent model development, and peer-reviewed validation will be essential as we scale these efforts to position ML as a cornerstone of embryo evaluation and selection.

.....

ACKNOWLEDGMENTS

The authors wish to disclose that they are shareholders in EmGenisys, the company responsible for the development of the embryo analysis software discussed in this study. Additionally, Cara Wells and Russell Killingsworth are recognized as inventors of the technology underlying this innovation. These financial interests may be perceived as potential conflicts of interest. The authors have disclosed these interests to the journal to ensure transparency and uphold the integrity of the research process.

AUTHOR'S CONTRIBUTIONS

MR and RK recorded videos of embryos utilized in this study. RK, MR, and CW labeled stage and grade data in this study. CH, RK, and CW developed the survey and sur-

veyed embryologists. BC trained, validated and tested machine learning models. CH, RK and CW were each involved in data analytics. CW prepared the manuscript, and all authors were involved in editing.

CORRESPONDING AUTHOR

Cara Wells - cara@emgenisys.com

COMPETING OF INTEREST

Authors have conflict of interest to declare. CW, RK, CH and MR own stock in EmGenisys, which owns all intellectual property associated with this work.

INFORMED CONSENT STATEMENT

All authors and institutions have confirmed this manuscript for publication

DATA AVAILABILITY STATEMENT

Additional data were collected in this study beyond embryo stage and grade, including pregnancy outcome, embryo sex, genetic information, animal registration number, farm identity, and other identifying details. Because many of the animals in this dataset are marketed for their breeding value, maintaining their confidentiality is essential. Furthermore, pregnancy outcome data will be used for future investigations and have not been anonymized. As a result, the full dataset is not publicly available. However, an anonymized version of the data may be obtained from the corresponding author upon reasonable request.

Published: July 23, 2025 CDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

REFERENCES

1. Betteridge KJ. A history of farm animal embryo transfer and some associated techniques. *Anim Reprod Sci.* 2003;79:203-244. doi:[10.1016/S0378-4320\(03\)00166-0](https://doi.org/10.1016/S0378-4320(03)00166-0)
2. Elsdon RP, Hasler JF, Seidel GE. Non-surgical recovery of bovine eggs. *Theriogenology.* 1976;6(5):523-532. doi:[10.1016/0093-691X\(76\)90120-5](https://doi.org/10.1016/0093-691X(76)90120-5)
3. Rowe RF, Del Campo MR, Eilts CL, French LR, Winch RP, Ginther OJ. A single cannula technique for nonsurgical collection of ova from cattle. *Theriogenology.* 1976;6(5):471-483. doi:[10.1016/0093-691X\(76\)90114-X](https://doi.org/10.1016/0093-691X(76)90114-X)
4. Drost M, Brand A, Aarts MH. A device for nonsurgical recovery of bovine embryos. *Theriogenology.* 1976;6(5):503-507. doi:[10.1016/0093-691X\(76\)90117-5](https://doi.org/10.1016/0093-691X(76)90117-5)
5. Viana JHM. Statistics of embryo production and transfer in domestic farm animals. *IETS Data Retrieval.* Published online 2023.
6. Bó GA, Maplettoft RJ. Evaluation and classification of bovine embryos. *Anim Reprod.* 2013;10(3):344-348.
7. Rabel RAC, Marchioreto PV, Bangert EA, Wilson K, Milner DJ, Wheeler MB. Pre-implantation bovine embryo evaluation—from optics to omics and beyond. *Animals (Basel).* 2023;13(13):2102. doi:[10.3390/ani13132102](https://doi.org/10.3390/ani13132102). PMID:37443900
8. Rocha JC, Passalia FJ, Matos FD, Maserati MP Jr, Alves MF, De Almeida TG, et al. Methods for assessing the quality of mammalian embryos. *JBRA Assist Reprod.* 2016;20(3):150-158.
9. Donnay I, Partridge RJ, Leese HJ. Can embryo metabolism be used for selecting bovine embryos before transfer? *Reprod Nutr Dev.* 1999;39:523-533.
10. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach.* 4th ed. Pearson; 2021.
11. McCarthy J, Minsky M, Papert S, Shannon C. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* Dartmouth College; 1956.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444. doi:[10.1038/nature14539](https://doi.org/10.1038/nature14539)
13. Hsu W. *Behind Deep Blue: Building the Computer That Defeated the World Chess Champion.* Princeton University Press; 2002.
14. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell.* 1986;8(6):679-698. doi:[10.1109/TPAMI.1986.4767851](https://doi.org/10.1109/TPAMI.1986.4767851)
15. Brown TB et al. Language models are few-shot learners. *Adv Neural Inf Process Syst.* Published online 2020.
16. Chen L et al. DeepDriving: Learning affordance for direct perception in autonomous driving. In: *Proc CVPR.* ; 2015. doi:[10.1109/ICCV.2015.312](https://doi.org/10.1109/ICCV.2015.312)
17. Belcic I, Stryker C. What is supervised learning? IBM. Accessed April 15, 2025. <https://www.ibm.com/think/topics/supervised-learning>
18. Salih M, Austin C, Warty RR, Tiktin C, Rolnik DL, Momeni M, et al. Embryo selection through artificial intelligence versus embryologists: a systematic review. *Hum Reprod Open.* 2023;2023(3). doi:[10.1093/hropen/hoad031](https://doi.org/10.1093/hropen/hoad031). PMID:37588797
19. Khosravi P, Kazemi E, Zhan Q, Malmsten JE, Toschi M, Zisimopoulos P, et al. Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *NPJ Digit Med.* 2019;2:21. doi:[10.1038/s41746-019-0096-y](https://doi.org/10.1038/s41746-019-0096-y). PMID:31304368
20. Fertility Bridge. AI IVF company introduces Pay-for-Baby model. April 17, 2024. <https://www.fertilitybridge.com/news-articles/ivf-pay-for-baby-model-ai-aiivf>
21. Ducharme J. How artificial intelligence could change the fertility world. *TME.* Published online January 31, 2019. <https://time.com/5492063/artificial-intelligence-fertility/>
22. Barash OO, Miller KA, Wozniak M, et al. Evaluation of the effect on ongoing pregnancy rate of using artificial intelligence for embryo prioritization: An interim analysis of a prospective randomized control trial [Abstract]. *Fertility and Sterility.* 2024;122(4):e347. doi:[10.1016/j.fertnstert.2024.08](https://doi.org/10.1016/j.fertnstert.2024.08)
23. Wang S, Chen L, Sun H. Interpretable artificial intelligence-assisted embryo selection improved single-blastocyst transfer outcomes. *Reprod Biomed Online.* 2023;47(6):103371. doi:[10.1016/j.rbmo.2023.103371](https://doi.org/10.1016/j.rbmo.2023.103371)

24. Weill Cornell Medicine. A fully automated AI-based system can assess IVF embryo quality.
25. VerMilyea M, Diakiw SM, Hall JMM, et al. Non-invasive, image-based genetics AI improves live birth outcomes — even for embryos classified as euploid by PGT-A [Poster presentation]. Pacific Coast Reproductive Society (PCRS) 2024 annual meeting. 2024. <https://www.lifewhisperer.com/wp-content/uploads/2024/05/PCRS-2024-Tex-VerMilyea.pdf>
26. Yuan H, Wei X, Lu Y, Yao Z, Liu W, Wu Y, et al. A multi-modal AI system for embryo assessment. *Patterns*. 2024;5(5):100100.
27. Presagen. Life Whisperer Genetics AI improves live birth outcomes.
28. Wells C, Hayden CB, Rea M, Killingsworth R. Advancing embryo evaluation: Generative AI in routine embryo transfer. conference presentation presented at: Annu Int Embryo Technol Soc Conf; January 18–22, 2025; Fort Worth, TX.
29. Rocha JC, Passalia FJ, Matos FD, Takahashi MB, Maserati MP Jr, Alves MF, et al. Automatized image processing of bovine blastocysts. *Sci Data*. 2017;4:170192. doi:[10.1038/sdata.2017.192](https://doi.org/10.1038/sdata.2017.192). PMID:29257125
30. Farin P, Britt J, Shaw D, Slenning B. Agreement among evaluators of bovine embryos. *Theriogenology*. 1995;44:339-349. doi:[10.1016/0093-691X\(95\)00189-F](https://doi.org/10.1016/0093-691X(95)00189-F)
31. Belcic I, Stryker C. What is supervised learning? IBM. Accessed April 15, 2025. <https://www.ibm.com/cloud/learn/supervised-learning>
32. Neethirajan S. The role of sensors, big data and machine learning in modern animal farming. *Sens Bio-Sens Res*. 2020;29:100367. doi:[10.1016/j.sbsr.2020.100367](https://doi.org/10.1016/j.sbsr.2020.100367)
33. Lu J. Will medical technology deskill doctors? *Int Educ Stud*. 2016;9:130.
34. Duran LD. Deskillling of medical professionals: an unintended consequence of AI implementation? *Giornale di Filosofia*. 2021;2:47-59.
35. Koplin JJ, Johnston MN, Webb ANS, Whittaker A, Mills C. Ethics of AI in embryo assessment. *Hum Reprod*. 2025;40(2):179-185. doi:[10.1093/humrep/deae264](https://doi.org/10.1093/humrep/deae264). PMID:39657965