OBJECTIVE

Analysis of clinical data suggests there are inherent errors in the classification of Day 5 blastocyst images used for developing viability prediction algorithms, where embryos that appear morphologically viable are wrongly classified as non-viable, based on a negative pregnancy outcome occurring due to external factors. A novel artificial intelligence technique (UDC)* was used to identify and remove misclassified, poor quality data to obtain a cleaned dataset to train and test an artificial intelligence (AI) model for viability prediction. This improves AI performance and reduces misleading clinical reporting of AI accuracy.

METHOD

Two methods were used to verify that embryos classified as non-viable can often be viable, with other non-morphological patient factors such as endometriosis contributing to non-pregnancy.

1. **Clinical Analysis**

   5,500 static 2D images of Day 5 blastocysts with associated clinical pregnancy outcomes were obtained from 15 clinics in 5 countries and assessed for morphology and patient fertility factors. Clinical analysis primarily considered patients under 35 years as they are likely to have a higher proportion of viable embryos, with patient factors the most likely cause contributing to failed pregnancy in this age bracket.

2. **AI Analysis**

a. A novel AI technique (UDC) which identifies misclassified clinical data was developed to identify morphologically viable embryos incorrectly classified as non-viable. To verify that misclassified non-viable data were correctly identified by the UDC, removing the misclassified (erroneous) data prior to AI training was expected to show an improvement in AI performance.

b. An AI trained model using a UDC cleaned dataset was tested for performance on i) an uncleaned and ii) a UDC-cleaned blind test dataset. The aim was to show that testing and reporting AI performance on a dataset that inherently comprises misclassified data can be misleading.

Patent filed: 54839AU-P

OVATION FERTILITY

Life Whisperer
A.I. Enhanced Fertility

IDENTIFYING INHERENT POOR QUALITY EMBRYO DATA USING ARTIFICIAL INTELLIGENCE TO IMPROVE AI PERFORMANCE AND CLINICAL REPORTING

ASRM 2020
P-100

TAKE A PICTURE TO
DOWNLOAD PRESENTATION

Texas Fertility Center

M. A. Dakka[3]*, M.D. VerMilyea[1,2,5,6], T. Nguyen[3], D. Perugini[3], S. M. Diakiw[3], K. Silverberg[1,2,6], J. Hall[3,7]*, M Perugini[3]#

[1]Ovation Fertility-Austin, TX; [2]Ovation Fertility – San Antonio, TX; [3]Presagen & Life Whisperer, Adelaide, Australia; [4]Ovation Fertility Genetics – Henderson, NV; [5]California Fertility Partners – Los Angeles, CA; [6]Texas Fertility Center- Austin, TX; [7]Australian Research Council Centre of Excellence for Nanoscale BioPhotonics, The University of Adelaide, Adelaide, Australia; *Equal First Author; #Presenting Author

RESULTS

1. Clinical Analysis

30% Increase in patient factors reported for patients under 35 who did not get pregnant (63.9%), compared with patients over 35 years (49.1%).

49% Of embryos in the <35-year age group did not lead to a pregnancy, despite **only 17% of embryos** being deemed non-viable by traditional morphological grading.

2. AI Analysis

Accuracy and sensitivity increase: When AI is **trained** using the UDC-cleaned dataset and **tested** on the non-cleaned blind test dataset.

	AI trained on non-cleaned dataset	AI trained on cleaned dataset
Overall accuracy	59.7%	61.1%
Accuracy for viable embryos (sensitivity)	76.8%	80.6%
Accuracy for non-viable embryos (specificity)	37.3%	35.4%

77% True AI overall performance as measured on the UDC **cleaned test dataset**, where patient factors are controlled for, compared to 61.1% on the non-cleaned dataset.

CONCLUSION

Results demonstrate that embryos classified as non-viable due to a negative pregnancy outcome, can sometimes be morphologically viable and misclassified for the purposes of predictive algorithms. The UDC is a unique technique that is effective at identifying these misclassified cases, which results in improved AI performance and enables the true reporting of AI performance when incorrect data are removed prior to AI training. This highlights key methodological flaws in reporting ultra high accuracies (above 90%) in examples of recent literature, when embryo viability data quality has not been adequately controlled for.