P-165 Using Artificial Intelligence to Classify Embryo Shape: An International Perspective

R. Hariharan[1], C. He[1], C. Hickman[1], J. Chambost[2], C. Jacques[2], M. Hentschke[3], B. Cunegatto[4], C. Dutra[5], A. Drakeley[6], Q. Zhan[7], R. Miller[8], G. Verheyen[9], M. Rosselot[10], S. Loubersac[10], K. Kelley[11]

[1]Apricity, AI Team, London, United Kingdom ;
[2]Apricity, AI Team, Paris, France ;
[3]Fertilitat, Gynaecology, Porto Alegre, Brazil ;
[4]Fertilitat, Embryology, Porto Alegre, Brazil ;
[5]Reproferty, Embryology, São José dos Campos, Brazil ;
[6]Hewitt Fertility Centre of Liverpool Women's Hospital, Obstetrics and Gynaecology, Liverpool, United Kingdom ;
[7]Weill Cornell Medicine, Obstetrics and Gynaecology, New York, U.S.A. ;
[8]Weill Cornell Medicine, Reproductive Medicine, New York, U.S.A. ;
[9]UZ Brussels, Reproductive Medicine, Jette, Belgium ;
[10]CHU de Nantes, Reproductive Medicine, Nantes, France ;
[11]POMA Fertility, Data Analytics, Kirkland, U.S.A.

Study question: Is a pre-trained machine learning algorithm able to accurately detect cellular arrangement in 4-cell embryos from a different continent?
Summary answer: Artificial Intelligence (AI) analysis of 4-cell embryo classification is transferable across clinics globally with 79% accuracy.
What is known already: Previous studies observing four-cell human embryo configurations have demonstrated that non-tetrahedral embryos (embryos in which cells make contact with fewer than 3 other cells) are associated with compromised blastulation and implantation potential. Previous research by this study group has indicated the efficacy of AI models in classification of tetrahedral and non-tetrahedral embryos with 87% accuracy, with a database comprising 2 clinics both from the same country (Brazil). This study aims to evaluate the transferability and robustness of this model on blind test data from a different country (France).
Study design, size, duration: The study was a retrospective cohort analysis in which 909 4-cell embryo images ("tetrahedral", n = 749; "non-tetrahedral", n = 160) were collected from 3 clinics (2 Brazilian, 1 French). All embryos were captured at the central focal plane using Embryoscope™ time-lapse incubators. The training data consisted solely of embryo images captured in Brazil (586 tetrahedral; 87 non-tetrahedral) and the test data consisted exclusively of embryo images captured in France (163 tetrahedral; 72 non-tetrahedral).
Participants/materials, setting, methods: The embryo images were labelled as either "tetrahedral" or "non-tetrahedral" at their respective clinics. Annotations were then validated by three operators. A ResNet-50 neural network model pretrained on ImageNet was fine-tuned on the training dataset to predict the correct annotation for each image. We used the cross entropy loss function and the RMSprop optimiser (lr = 1e-5). Simple data augmentations (flips and rotations) were used during the training process to help counteract class imbalances.

Main results and the role of chance: Our model was capable of classifying embryos in the blind French test set with 79% accuracy when trained with the Brazilian data. The model had sensitivity of 91% and 51% for tetrahedral and non-tetrahedral embryos respectively; precision was 81% and 73%; F1 score was 86% and 60%; and AUC was 0.61 and 0.64. This represents a 10% decrease in accuracy compared to when the model both trained and tested on different data from the same clinics.

Limitations, reasons for caution: Although strict inclusion and exclusion criteria were used, inter-operator variability may affect the pre-processing stage of the algorithm. Moreover, as only one focal plane was used, ambiguous cases were interpoloated and further annotated. Analysing embryos at multiple focal planes may prove crucial in improving the accuracy of the model.

Wider implications of the findings: Though the use of machine learning models in the analysis of embryo imagery has grown in recent years, there has been concern over their robustness and transferability. While previous results have demonstrated the utility of locally-trained models, our results highlight the potential for models to be implemented across different clinics.

Trial registration number: Not Applicable