An artificial intelligence model that was trained on pregnancy outcomes for embryo viability assessment is highly correlated with Gardner score

Sonya Diakiw[1], M. VerMilyea[2,3], J .M.M. Hall[1,4], K. Sorby[5], T. Nguyen[1], M.A. Dakka[1], D. Perugini[1], M. Perugini[1]

[1]Presagen, Life Whisperer, Adelaide, Australia
[2]Ovation Fertility, Laboratory, Austin, U.S.A.
[3]Texas Fertility Center, IVF Laboratory, Austin, U.S.A.
[4]Australia/Australian Research Council Centre of Excellence for Nanoscale BioPhotonics, The University of Adelaide, Adelaide, Australia
[5]No.1 Fertility, Melbourne, Australia

Life Whisperer
A.I. Enhanced Fertility

OVATION FERTILITY

Nº 1 FERTILITY

eshre
www.eshre.eu
European Society of Human Reproduction and Embryology

Introduction

Correlation between AI and Gardner score for embryo viability assessment?

Life Whisperer's AI was trained to evaluate embryo viability, i.e. the likelihood of an embryo leading to clinical pregnancy (foetal heartbeat) - *VerMilyea et al, 2020 (Hum Rep)*

Gardner score is a common visual inspection method for embryo selection based on morphology

This study addresses three questions:

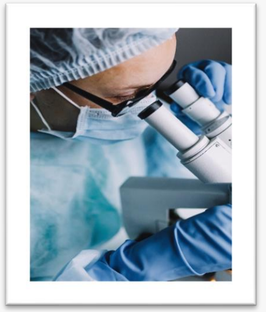
Do Life Whisperer AI scores correlate with known visible features of embryo development (Gardner score)?

Do the AI scores correlate with pregnancy outcomes according to Gardner score, as might be expected?

How does the AI compare with Gardner score in terms of ability to predict pregnancy outcome?

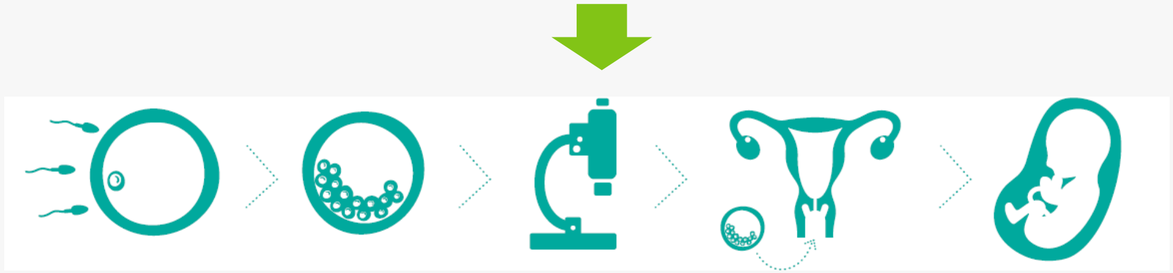
Embryo selection

Gardner Score



VS

Artificial Intelligence



Study Design & Methods

Prospectively collected dataset from No. 1 Fertility

Study conducted on 2,162 prospectively collected single Day 5 embryo images (Embryoscope) with associated Gardner scores and AI viability scores

To avoid bias embryologists manually graded each embryo using the Gardner method first, then obtained the AI score

AI viability scores are between 0 (least likely to lead to pregnancy, non-viable) and 10 (most likely to lead to pregnancy, viable)

The AI was trained and validated on 3,651 Day 5 embryo images with pregnancy outcomes from multiple IVF laboratories across 5 countries, but was not trained on data used in this study

Characteristic	Value
Number of images of Day 5 embryos with Gardner score	2,162
Number of patients	787
Average embryo cohort size (range)	2.3 embryos (1-18 embryos)
Average patient age (range)	36.2 years (24-49 years)
Dates collected	Nov 2019 – Feb 2021
Expansion grades: [1]	
1 = Early or very early blastocyst	15 (0.7%)
2 = Blastocyst	85 (3.9%)
3 = Full blastocyst	486 (22.5%)
4 = Expanded blastocyst	466 (21.6%)
5 = Hatching blastocyst	1,097 (50.7%)
6 = Hatched blastocyst	13 (0.6%)
Number of images with fetal heartbeat outcomes [2]	479 (22.2%)
Clinical pregnancy outcomes:	
Successful pregnancies (viable embryos)	220 (45.9%)
Unsuccessful pregnancies (non-viable embryos)	259 (54.1%)

1 Expansion grades were based on the 6 stages of blastocyst expansion defined by Gardner and Schoolcraft.
2 Only single embryo transfers were included.

AI scores correlate with Gardner score

Average AI score significantly correlated with the three components of the Gardner score, increasing with advancing blastocyst developmental stage and increasing ICM/TE quality

Expansion Grade

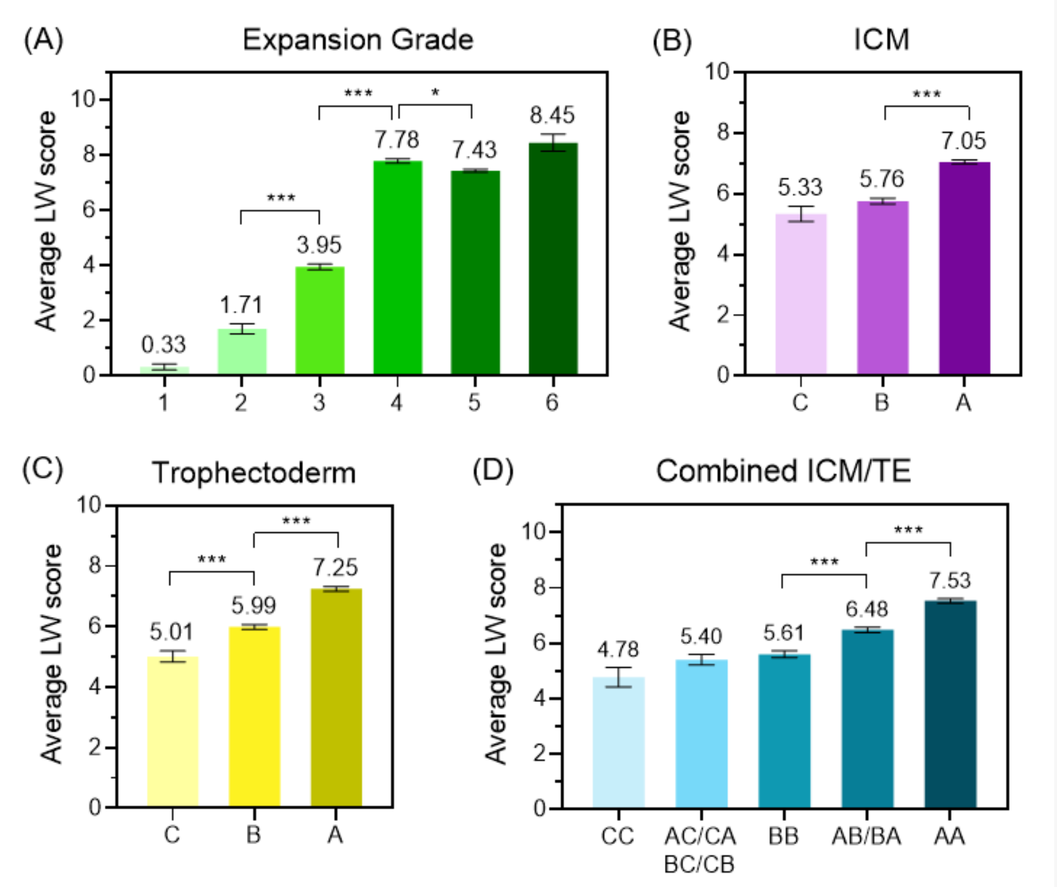
Grades of ≥3 generally considered suitable for transfer

Grade 3 had lower average score than 4 to 6, consistent with some published research that grade 3 (full blastocysts) may have lower clinical pregnancy rates than more advanced expansion stages

Inner Cell Mass (ICM) & Trophectoderm (TE)

AI score showed stronger linear correlation with TE than ICM, consistent with studies that suggest TE may be more important than ICM in determining likelihood of implantation

Average AI score correlated with increasing combined ICM/TE grade



Groups were compared using ordinary one-way ANOVA with Tukey's multiple comparisons post-test, and trends were evaluated using ordinary one-way ANOVA with test for linear trend between column mean in left-to-right column order.

AI scores correlate with pregnancy rate

Clinical pregnancy rate also correlated with the three components of the Gardner score

Expansion Grade

Grade 3 had lower pregnancy rate than then 4 and 6

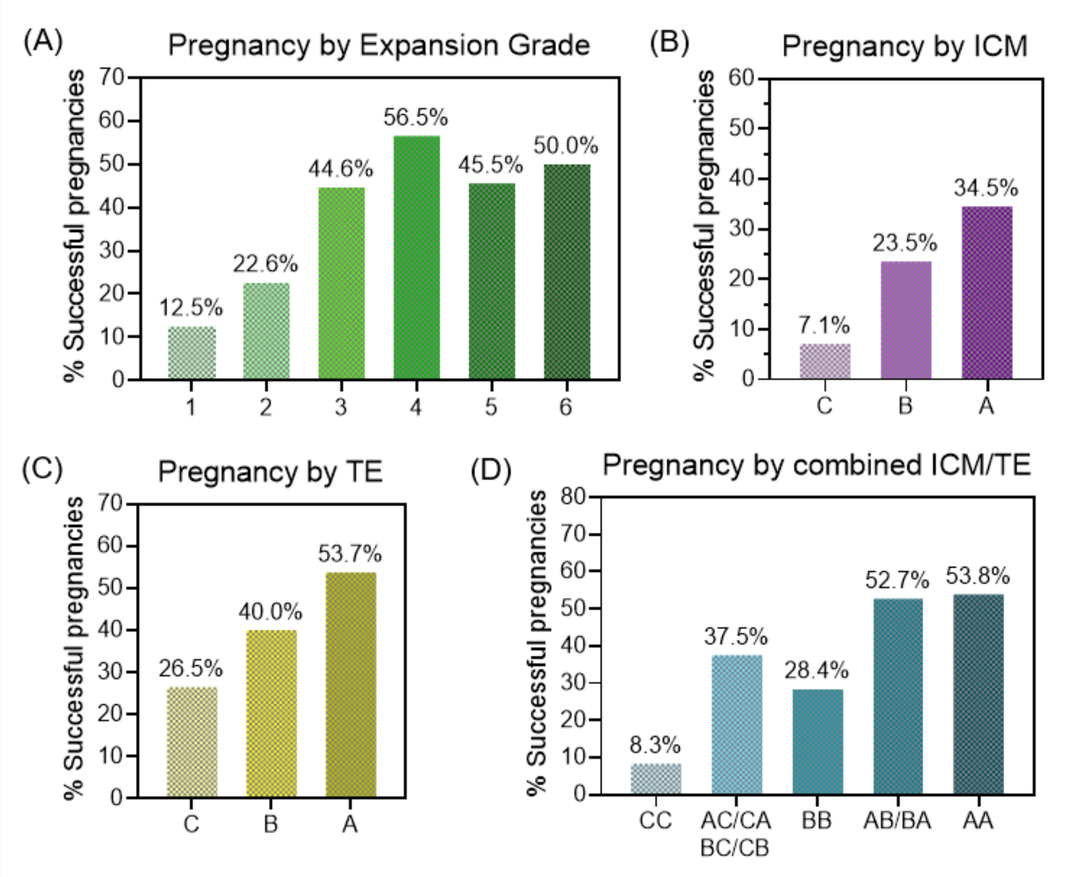
Reduced pregnancy rate for grade 5 embryos (in alignment with reduced average AI score)

Inner Cell Mass (ICM) & Trophectoderm (TE)

Linear correlation with ICM and TE, including difference between ICM grades C and B

Supports quality of the dataset, and correlation of AI score with pregnancy rate

Does not take into consideration patient-related medical factors that may influence outcome of a viable embryo



Subset of 479 embryo images had pregnancy outcomes out of the overall dataset of 2,162 images (22%)

Thresholds for embryo viability

What are the challenges of defining a ‘good’ versus ‘poor’ embryo for evaluating prediction algorithms?

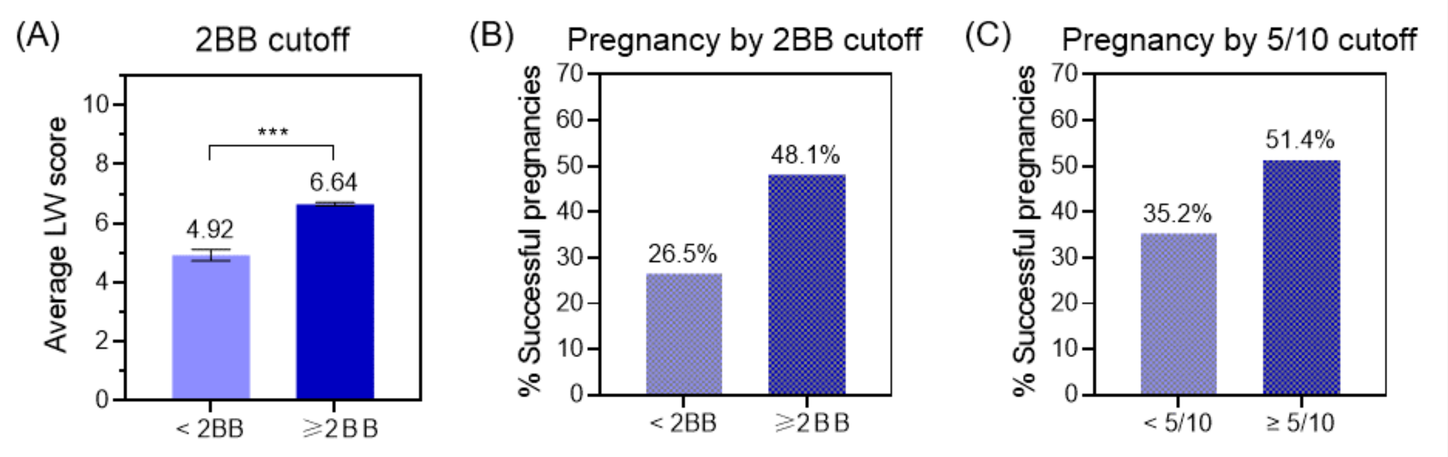
Limitation of Gardner score = not linear scale

No standard for defining what is likely to be a viable embryo → 3AA, 3BB, 2BB?

The 2BB threshold for this study was chosen based on literature (Munné et al 2019) and discussion with experienced embryologists

Significant difference in average AI score, verified 2BB threshold using pregnancy data

5/10 threshold for AI also verified using pregnancy data



Groups were compared using an unpaired 2-tailed student’s t-test

Predictive power of AI versus Gardner score

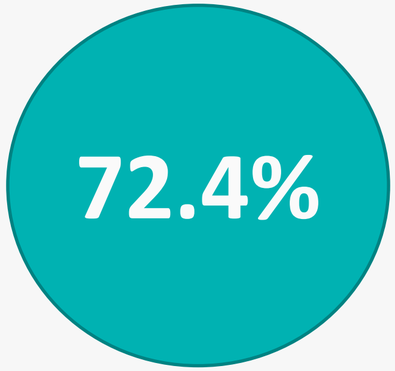
AI predicts Gardner score, but is better at predicting pregnancy than the Gardner score itself

AI for predicting Gardner score

Accuracy of 72.4% for predicting embryos of ≥2BB, showing there is a significant overlap in detection of known features of blastocyst morphology identified by the AI and Gardner methods.

AI versus Gardner score for predicting pregnancy

Accuracy of AI was 10.3% higher than the Gardner for predicting pregnancy, showing that the AI is not just detecting the same features as the Gardner system.



AI accuracy for predicting Gardner Grade

AI score ≥5/10 was used to predict embryos with Gardner ≥2BB

AI accuracy **improvement** over Gardner for predicting pregnancy outcome

Viable embryo prediction: AI score ≥5/10, Gardner Grade ≥2BB

Feature identification

Gardner score

ICM (grade A)

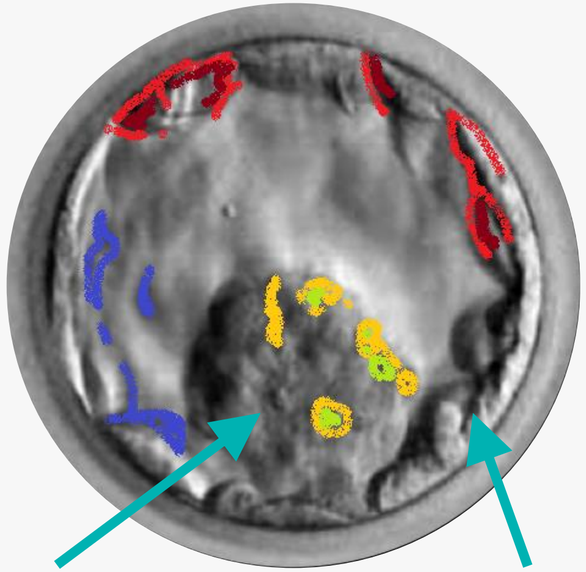
TE (grade B)

Expansion grade 4
(expanded blastocyst)

The AI is identifying additional morphological features that are not captured using the Gardner scoring method, but which are directly associated with pregnancy outcome.

→ What are these features? Future research.

AI algorithm?



ICM (grade A)

TE (grade B)

Expansion grade 4
(expanded blastocyst)

Conclusions

AI provides additional information over Gardner about pregnancy outcomes

- Correlation between AI and known features of embryo viability (Gardner score) substantiates the use of this AI for embryo assessment
- AI has also shown to correlate with embryo ploidy status, further supporting use for embryo assessment – refer to our ESHRE 2021 poster P-228
- Results support the use of AI to provide additional information regarding embryo viability and pregnancy outcome over and above Gardner score
- Limitations
  - This correlative study may also require additional confirmatory studies on independent datasets
  - Need to consider limitations of using a binary threshold for evaluating performance of an AI
- In addition to improved accuracy of the AI for predicting pregnancy, there are also many intangible benefits of AI in the IVF laboratory including standardisation of scoring, objectivity of assessment, scoring efficiency, and transparency for the patient

eshre