Embryologists' predictions do not improve an artificial intelligence (AI) tool's prediction of blastocyst development from mature (metaphase II, MII) oocytes when augmented in ensemble models

Natalie Mercuri, Sergey Corsac, Jullin Fjeldstad, Nadia Siddique, Isabel Puerta Vega, Dan Nayot

Objective

To assess the performance of ensemble prediction models combining embryologists' weighted predictions with an AI tool's (VIOLET, Future Fertility) prediction of blastocyst development from images of MII oocytes.

Methods

- VIOLET analyzed 300 static images of MII oocytes to predict blastocyst development
- 17 embryologists (3 clinic groups) asked to predict blastocyst development (best judgment) from same images and score their confidence (1-3) in each prediction
- Weighted probability for each oocyte was calculated and utilized to create 2 ensemble prediction models combining VIOLET's and embryologists' probabilities of blastocyst development
- **Ensemble 1** utilized lambda value (0, 0.25, 0.5, 0.75, or 1) — higher lambda places more weight on VIOLET's probability in the prediction
- **Ensemble 2** utilized VIOLET's confidence as thresholds (10%, 30%, 50%, 70%, 90%). If VIOLET's confidence exceeded the threshold, its prediction was used, otherwise embryologists' weighted prediction was used
- Accuracy, specificity, sensitivity, and AUC were calculated to assess performance of both ensembles.
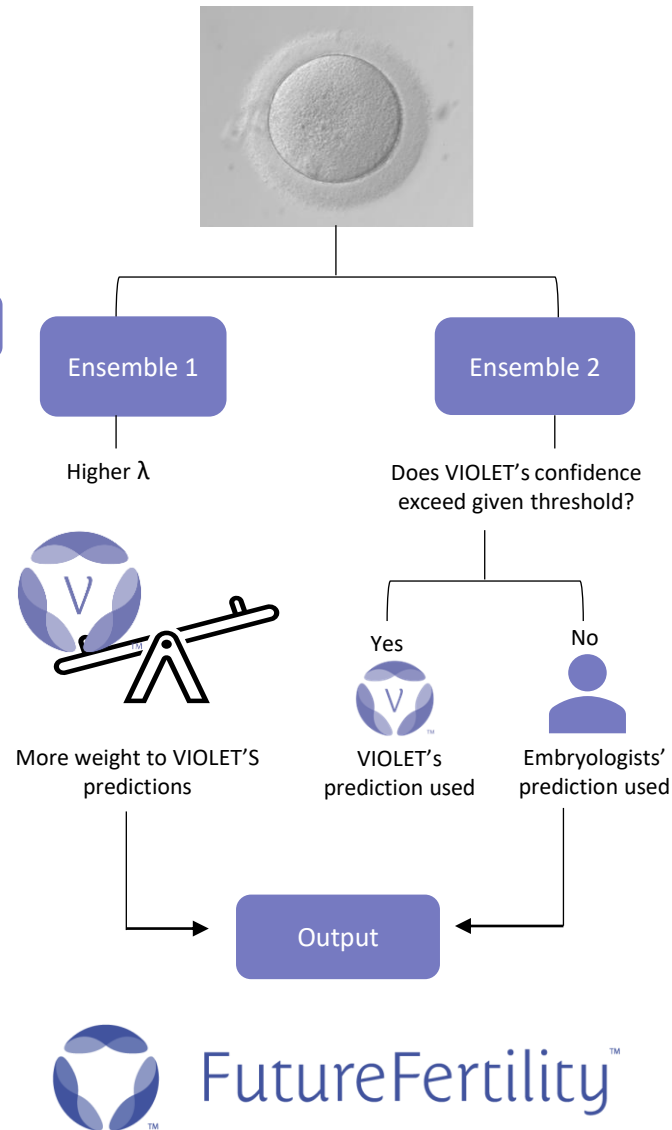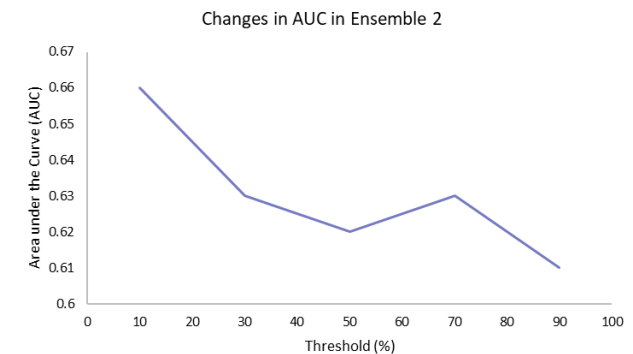

Results

Ensemble 1 displayed stepwise increase in prediction accuracy (0.54–0.61), AUC (0.61–0.66), and specificity (0.32–0.67), and a decrease in sensitivity (0.9–0.53), as lambda value increased from 0 to 1 (more weight on VIOLET's probability).
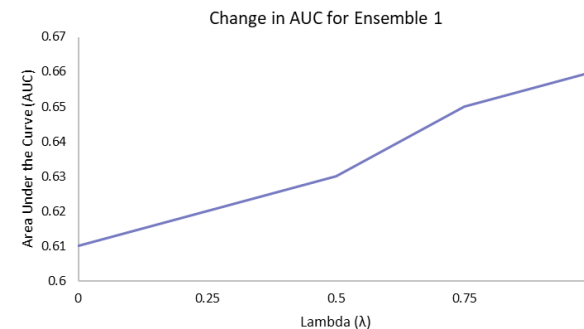
Ensemble 2 displayed decrease in prediction accuracy (0.62–0.56), AUC (0.66–0.61), and specificity (0.63–0.33) as threshold increased from 10% to 90% (increasing utilization of embryologist prediction); yet sensitivity increased (0.61-0.89).

λ values of 0, 0.25, 0.5, 0.75, & 1.

Confidence thresholds of 10%, 30%, 50%, 70% & 90%.

Conclusions

VIOLET's predictions result in a balanced and higher overall model performance, suggesting VIOLET can account for relevant oocyte features detectable by the human eye, while extracting additional information that is imperceptible – providing a consistent, efficient and more accurate assessment.

Conflict of Interest

Authors are employees of Future Fertility

Funding

No external sources of funding were received.