

Comparison of manual and automated ovarian follicle measurement methods

Piotr Wygocki^{1,2}, Mateusz Ulfig¹, Andrzej Zapala^{1,2}, Marcin Zieleni¹, Krystian Zieliński^{3,4}, Natalia Gajewska³, Damian Drzyzga³, Marcin Wrochna^{1,2}
 1 MIM Solutions | 2 University of Warsaw | 3 INVICTA Research and Development Center | 4 Gdańsk University of Technology || Poland

PURPOSE & OBJECTIVES

When measuring follicles via transvaginal ultrasound, published recommendations are to take the mean of two perpendicular diameters. Nonetheless, follicle boundaries are often ambiguous and differently assessed by different sonographers and methods. It is also common to use only a single diameter to speed up the process, especially when there are many follicles.

In this study, we tested whether automating follicle detection with artificial intelligence may improve reliability and consistency of measurements. FOLLISCAN is a deep learning AI algorithm which provides exact outlines of every follicle in a full automated fashion.

MATERIAL & METHODS

For human expert to AI model comparison

- ◆ 27 patients undergoing stimulation for IVF
- ◆ 102 videos (cine-loops of ovary sweep)
- ◆ 1869 follicles
- ◆ Processed with FOLLISCAN
- ◆ Independently annotated by 3 experts
- ◆ Consensus annotations by the experts

For comparing diameter methods (single, double diameters, or follicle area)

- ◆ 1000 videos
- ◆ 12167 follicles
- ◆ Processed with FOLLISCAN

RESULTS

Follicle detection

	On follicle ≥8mm (N=702)		On all follicles (N=1869)	
	Precision (95% CI)	Recall (95% CI)	Precision (95% CI)	Recall (95% CI)
AI Model	95.6% (93.9-97.1)	86.6% (82.7-89.6)	89.7% (87.8-91.5)	69.6% (66.4-72.6)
Expert A	98.9% (97.9-99.6)	78.1% (67.8-85.0)	93.6% (89.7-95.6)	71.7% (65.1-76.7)
Expert B	97.9% (94.1-99.1)	91.2% (86.2-93.8)	94.4% (91.7-96.0)	85.2% (82.4-87.3)
Expert C	91.5% (86.9-94.4)	86.0% (81.2-89.3)	89.7% (86.2-92.1)	71.3% (68.4-73.9)

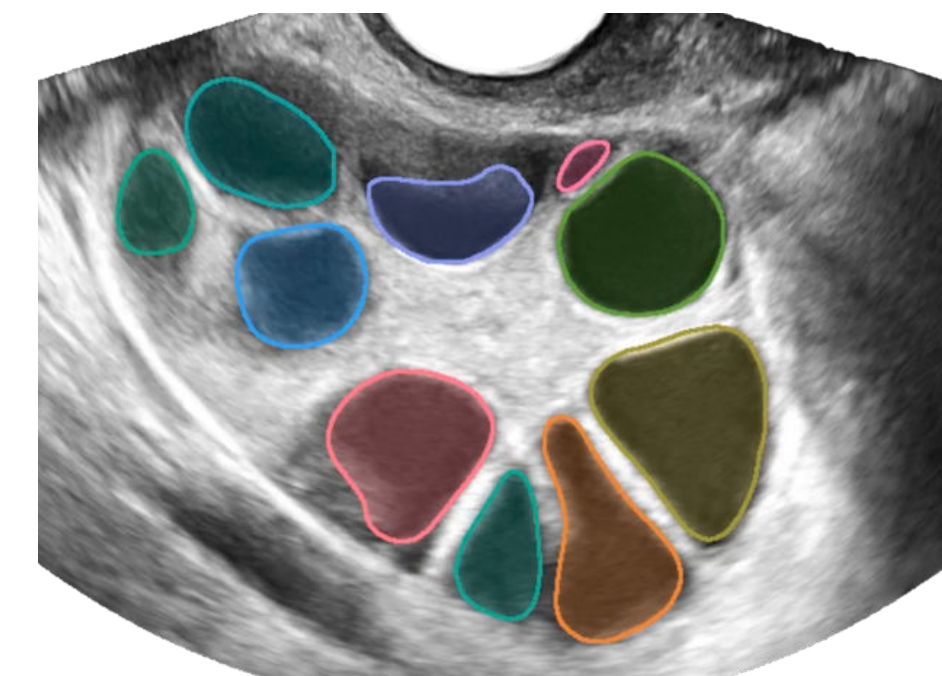
Inter-observer agreement

Mean absolute error on **833** follicles correctly found by all annotators

Expert vs expert (avg)	A vs B	A vs C	B vs C
0.83 mm	0.97 mm	0.91 mm	0.61 mm
Model vs expert (avg)	Model vs A	Model vs B	Model vs C
0.74 mm	0.93 mm	0.65 mm	0.64 mm

CONCLUSIONS

- ◆ **Speeds-up the process by 2.7x, average time per ovary:**
 - 80 s for reviewing and correcting automatic annotations
 - 220 s for manual annotation
- ◆ **Performance of AI model similar to human experts** (including difficult images of poor quality)
- ◆ **Standardizes measurements across a clinic, consistent and repeatable**
- ◆ Antral follicle count (AFC) highly correlated with consensus: Pearson's correlation **0.95** for fully automated, **0.78-0.97** for experts
- ◆ Volume- and area-based measurement reflect follicle shape
- ◆ High number of omissions observed in clinic's manual reports: Pearson's correlation **0.74** for archival AFC numbers provided by physicians, Mean absolute percentage error: **42.6%**
- ◆ Simplifying the process with automation is essential for affordably fast, yet reliable results with 2D or 3D probes



Methods of measuring diameters



Single diameter: 10.6 mm
 Double diameter mean measurement (12167 follicles): 8.8 mm
 Area equivalent: 8.5 mm

State-of-the-Art
AI-Enhanced Follicular Monitoring
2.7X Faster & More Reliable