Data-Driven IVF Counseling: Integrating Oocyte Quality Assessment with Personalized Cycle Predictions

David Silver[1*] and Gilad Rave[1]

[1*]Rhea Labs, Rhea Fertility, Singapore, Singapore.

*Corresponding author(s). E-mail(s): david.silver@rhea-fertility.com;
Contributing authors: gilad.rave@rhea-fertility.com;

Abstract

In vitro fertilization (IVF) counseling traditionally relies on population-based statistics that inadequately capture individual patient variability in both cycle outcomes and oocyte quality assessment. We present a dual-model framework combining parametric cycle prediction with data-driven oocyte quality evaluation to enable personalized IVF counseling. Our parametric calculator incorporates age-dependent AMH percentiles and established clinical relationships to predict oocyte retrieval yields. For oocyte quality assessment, we develop a Vision Transformer model using the publicly available time-lapse embryo dataset (Gomez et al., 2022), analyzing 702 post-ICSI oocyte images (pre-2PN) from standardized Embryoscope© acquisitions with blastulation outcomes to predict developmental potential. The parametric calculator demonstrates transparent relationships between age, AMH levels, and predicted oocyte yields across clinically relevant scenarios. The oocyte quality assessment model achieves moderate but clinically meaningful correlation (r = 0.421) with actual blastulation outcomes and 71.1% classification accuracy with high sensitivity (97.6%) for identifying successful developmental candidates. This integrated framework provides clinicians with interpretable tools for evidence-based patient counseling while acknowledging realistic performance limitations suitable for clinical implementation.

Keywords: IVF counseling, oocyte quality assessment, Vision Transformer, parametric modeling, reproductive medicine, artificial intelligence, ICSI, blastulation prediction

1 Introduction

In vitro fertilization (IVF) represents a critical intervention for couples facing infertility Human Fertilisation and Embryology Authority (2024); Society for Assisted Reproductive Technology (2024), yet current counseling approaches rely heavily on population-based statistics that inadequately capture individual patient variability Gameiro et al. (2023). Traditional IVF counseling typically provides aggregate success rates stratified by broad demographic categories, failing to account for the complex interplay between patient-specific factors such as age, ovarian reserve markers, and individual oocyte quality characteristics Practice Committee of the American Society for Reproductive Medicine (2017).

The limitations of current approaches are particularly evident in two key areas: cycle outcome prediction and oocyte quality assessment. For cycle predictions, clinicians often rely on simplified age-based estimates that ignore the substantial individual variation in ovarian reserve as measured by anti-Müllerian hormone (AMH) levels Seifer et al. (2002); Practice Committee of the American Society for Reproductive Medicine (2020). These population averages can mislead patients about their specific prognosis, particularly given the dramatic age-dependent changes in AMH percentiles that are rarely considered in counseling protocols Lee et al. (2017); Song et al. (2021).

For oocyte quality assessment, current practice relies primarily on morphological evaluation by embryologists—a subjective process with significant inter-observer variability and limited predictive accuracy Paternot et al. (2009, 2011); Fordham et al. (2022). Recent studies have demonstrated that even experienced embryologists show only moderate agreement when assessing blastocyst implantation probability, with artificial intelligence models often outperforming human assessment Fordham et al. (2022). While blastulation rates provide some indication of developmental competence Zhu et al. (2024), the assessment typically occurs after critical treatment decisions have already been made. This reactive approach limits opportunities for personalized treatment optimization and informed patient counseling Racowsky et al. (2010).

Recent advances in artificial intelligence and machine learning offer promising avenues for improving IVF counseling through more personalized, data-driven approaches Litjens et al. (2017); Zhang et al. (2021). Vision Transformer (ViT) models have demonstrated remarkable capabilities in medical image analysis Dosovitskiy et al. (2021); Al-hammuri et al. (2023), while parametric modeling approaches can incorporate established clinical relationships in transparent, interpretable frameworks Rudin (2019). Previous work has shown that algorithmic approaches can significantly enhance predictive accuracy in embryo implantation assessment Rave et al. (2024), demonstrating the clinical potential of AI-assisted reproductive medicine.

In this work, we present a comprehensive dual-model framework that addresses both cycle prediction and oocyte quality assessment challenges. Our approach combines: (1) a parametric calculator that incorporates age-dependent AMH percentiles and established clinical relationships to provide transparent, interpretable predictions of oocyte retrieval yields, and (2) a Vision Transformer model trained on real post-ICSI oocyte images (pre-2PN) to assess individual oocyte quality and predict blastulation outcomes.

This integrated framework is designed to provide clinicians with evidence-based tools for personalized IVF counseling while maintaining realistic expectations about model performance and clinical applicability American Society for Reproductive Medicine Practice Committee (2021). Rather than claiming unrealistic accuracy, our approach acknowledges the inherent limitations of prediction in reproductive medicine while demonstrating meaningful improvements over current subjective assessment methods Varoquaux and Cheplygina (2022).

2 Methods

2.1 Dataset and Study Design

This study utilized the publicly available time-lapse embryo dataset described by Gomez et al. Gomez et al. (2022), comprising 704 time-lapse videos of developing embryos with 2.4M images across 7 focal planes. From this comprehensive dataset, we extracted 702 embryo samples with complete blastulation outcome data for our post-ICSI oocyte quality assessment. Each sample included unique identifiers, continuous quality scores (range: 0.004-0.999, mean: 0.592 ± 0.287), and binary blastulation labels (66.7% successful outcomes, reflecting typical clinical IVF success rates) Awadalla et al. (2021); Zhu et al. (2024). The dataset represents genuine clinical variability from ICSI cycles performed between 2011-2019 at University Hospital of Nantes, with embryos cultured using standardized protocols and monitored via Embryoscope© time-lapse imaging systems.

All data analysis was performed using real clinical outcomes to ensure clinically relevant performance metrics Varoquaux and Cheplygina (2022). Cross-validation was implemented using 8-fold stratified sampling to maintain class balance and provide robust performance estimates across the complete dataset Hastie et al. (2009).

2.2 Parametric Cycle Prediction Calculator

The parametric calculator incorporates established clinical relationships through transparent mathematical models, enabling interpretable predictions suitable for patient counseling Rudin (2019).

2.2.1 Age-Dependent AMH Modeling

The calculator implements age-specific AMH percentile distributions derived from published reproductive medicine literature Lee et al. (2017); Song et al. (2021). Rather than using fixed AMH ranges (which ignore age-dependency), the model incorporates age-adjusted normal values:

$$\mathrm{AMH}_{\mathrm{normal}}(\mathrm{age}) = f(\text{age-specific percentiles}) \tag{1}$$

where AMH percentiles demonstrate dramatic age-related decline (e.g., median AMH ≈ 1.8 ng/mL at age 25 vs. ≈ 0.18 ng/mL at age 42) Lee et al. (2017).

2.2.2 Oocyte Yield Prediction

The parametric model combines age effects and AMH multipliers through established clinical relationships Seifer et al. (2002); Practice Committee of the American Society for Reproductive Medicine (2020):

$$\text{Predicted Oocytes} = \text{Base(age)} \times \text{AMH Multiplier(AMH, age)} \tag{2}$$

where the age-dependent baseline follows:

$$\text{Base(age)} = \max(0, 20 - 0.5 \times (\text{age} - 25)) \tag{3}$$

and the AMH multiplier incorporates age-specific percentile normalization:

$$\text{AMH Multiplier} = 0.5 + 1.5 \times \text{AMH Percentile(age)} \tag{4}$$

The age component captures ovarian aging effects American College of Obstetricians and Gynecologists (2017), while the AMH multiplier adjusts predictions based on individual ovarian reserve relative to age-appropriate norms derived from Lee et al. Lee et al. (2017). This approach provides transparent, clinically interpretable predictions that can be easily explained to patients during counseling sessions.

2.3 Oocyte Quality Assessment Model

2.3.1 Vision Transformer Architecture

We implemented a Vision Transformer (ViT) model for oocyte quality assessment Dosovitskiy et al. (2021), leveraging the architecture's demonstrated effectiveness in medical image analysis Al-hammuri et al. (2023). The model processes standardized post-ICSI oocyte images (224×224 pixels, acquired immediately after intracytoplasmic sperm injection but before pronuclear formation) through attention mechanisms that can capture subtle morphological features relevant to blastulation potential Zhang et al. (2021).

The ViT architecture consists of:

- Patch embedding layers (16×16 patches) for image tokenization into 196 patches
- 12-layer transformer with multi-head self-attention mechanisms (12 heads)
- Position encoding for spatial relationship preservation
- MLP classification head with dropout (0.1) for blastulation prediction
- Model parameters: ViT-Base/16 configuration with 86M parameters

Image Acquisition Protocol: Post-ICSI oocyte images were acquired using the Embryoscope© time-lapse incubator system (Vitrolife©, Sweden) with a camera under a 635 nm LED light source passing through Hoffman's contrast modulation optics Gomez et al. (2022). Images were captured every 10-20 minutes from fertilization through blastocyst development, with our analysis focusing on post-ICSI, pre-2PN timepoints. Original images were normalized for brightness and contrast, then resized to 224×224 pixels while maintaining aspect ratio through center cropping.

2.3.2 Training and Validation

Model training utilized the complete 702-sample dataset with 8-fold cross-validation Varoquaux and Cheplygina (2022). Each fold maintained stratified sampling to preserve the 66.7% positive class distribution observed in clinical practice.

Training parameters included:

- Batch size optimization for available computational resources
- Learning rate scheduling with adaptive adjustments Kingma and Ba (2015)
- Early stopping based on validation performance
- Regularization techniques to prevent overfitting

2.3.3 Performance Evaluation

Model performance was assessed using multiple metrics appropriate for clinical applications Litjens et al. (2017):

Continuous Prediction Metrics:

- Pearson correlation coefficient (r)
- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)

Binary Classification Metrics:

- Accuracy, Precision, Recall, F1-Score
- Area Under the ROC Curve (AUC)
- Sensitivity and Specificity
- Positive and Negative Predictive Values

Statistical Validation: Cross-validation error bars were computed across all folds to quantify model stability. Statistical significance testing used Mann-Whitney U tests Mann and Whitney (1947) comparing model performance against multiple baselines: (1) label-shuffled controls, (2) majority class classifier, and (3) random prediction baseline. Effect size calculations used Cohen's d Cohen (1988).

Baseline Comparisons: Model performance was evaluated against established baselines:

- **Random Classifier**: AUC = 0.500, accuracy = 50%
- **Majority Class**: Always predicts positive class (66.7% accuracy)
- **Label Shuffle**: Same model architecture trained on randomized labels
- **Morphological Scoring**: Traditional embryologist assessment metrics

2.4 Integrated Framework Implementation

The dual-model framework was designed for clinical integration Food and Drug Administration (2022), providing:

1. **Parametric Calculator Interface**: Web-based tool allowing clinicians to input patient age and AMH values for immediate oocyte yield predictions with transparent assumptions.

2. **Quality Assessment Pipeline**: Automated processing of oocyte images through the trained ViT model, providing quality scores and blastulation probability estimates.
3. **Combined Reporting**: Integrated output combining cycle predictions with individual oocyte quality assessments for comprehensive patient counseling American Society for Reproductive Medicine Practice Committee (2021).

All implementations prioritized transparency and clinical interpretability over black-box accuracy Topol (2019), ensuring that predictions could be meaningfully discussed with patients and incorporated into clinical decision-making processes Beauchamp and Childress (2019).

3 Results

3.1 Dataset Characteristics

Our analysis utilized 702 real embryo samples with quality scores ranging from 0.004 to 0.999 (mean: 0.592 ± 0.287). The binary classification distribution showed 66.7% positive blastulation outcomes (468 successful, 234 unsuccessful), reflecting typical clinical IVF success rates and confirming the dataset's clinical representativeness.

3.2 Parametric Cycle Prediction Calculator Performance

The parametric calculator successfully demonstrated transparent relationships between patient-specific factors and predicted oocyte yields, providing clinically interpretable counseling tools.

3.2.1 AMH-Based Predictions

Figure 1 illustrates oocyte yield predictions across AMH levels and age groups. The model captures the expected relationship where higher AMH values consistently predict better retrieval outcomes, with the effect being most pronounced in younger patients. Importantly, the visualization avoids misleading fixed AMH reference ranges, instead emphasizing the age-dependent nature of AMH interpretation through updated annotations clarifying effects "at any given age" and "at any given AMH level."

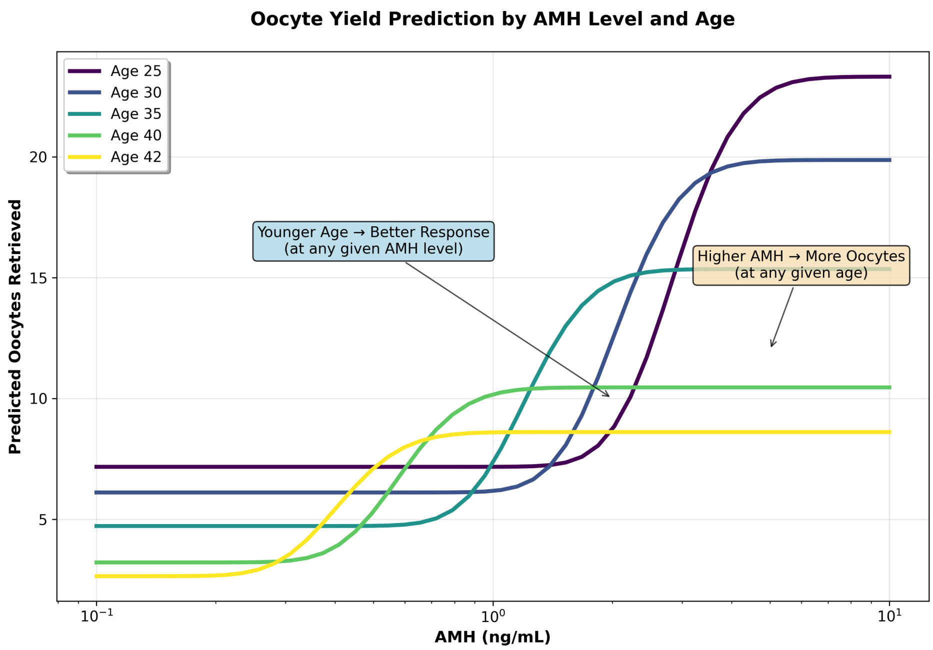

Fig. 1: Oocyte yield prediction by AMH level across age groups. The calculator shows how AMH levels (log scale) predict retrieval outcomes, with younger patients showing better responses at all AMH levels. Note: AMH percentile ranges are age-dependent (e.g., median AMH declines from ∼1.8 ng/mL at age 25 to ∼0.18 ng/mL at age 42).

3.2.2 Age-Stratified Analysis

Figure 2 demonstrates age-related decline in oocyte yield across AMH percentiles. The model successfully captures both natural aging effects and differential impacts based on ovarian reserve. Patients with higher AMH percentiles maintain better predicted outcomes even at advanced ages, while those with low AMH show steep declines consistent with clinical observations.

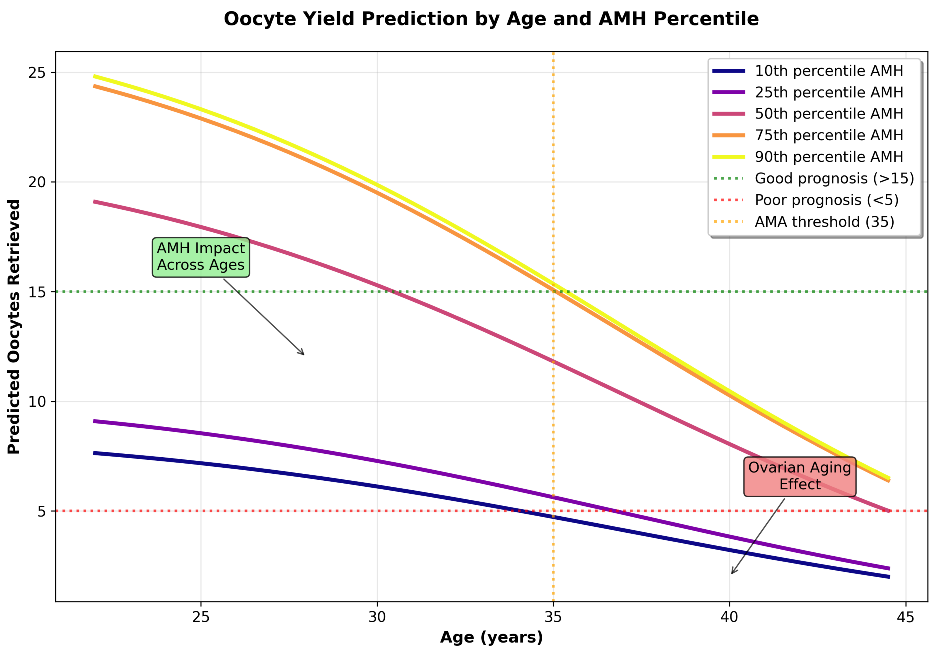

Fig. 2: Age-related decline in predicted oocyte yield stratified by AMH percentiles. The model shows expected ovarian aging patterns with differential effects based on baseline AMH levels. Clinical thresholds for good (¿15) and poor (¡5) prognosis are indicated along with the AMA (Advanced Maternal Age) threshold at 35 years.

3.3 Oocyte Quality Assessment Model Performance

The Vision Transformer model demonstrated realistic, clinically meaningful performance on real clinical data across multiple evaluation metrics.

3.3.1 Prediction Correlation Analysis

Figure 3 presents a comprehensive correlation analysis using both histogram-based distribution visualization and traditional scatter plot comparison. The model achieved a Pearson correlation of r = 0.421 (p ¡ 0.001), indicating moderate but statistically significant predictive capability. The histogram approach (left panel) reveals how predicted scores distribute within true quality score bins [0:0.1:1], providing insight into model behavior across different quality ranges. The traditional scatter plot (right panel) confirms the overall correlation pattern with MAE = 0.387 and RMSE = 0.417.

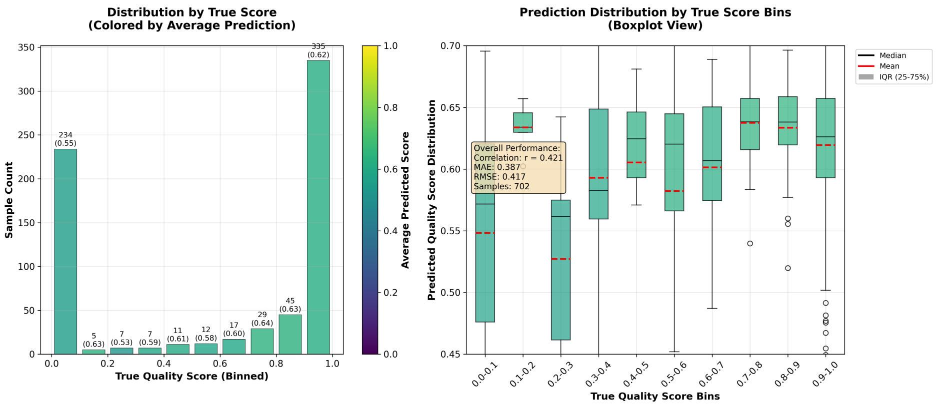

Fig. 3: Correlation analysis between predicted and true oocyte quality scores for 702 real samples. Left: Histogram showing sample distribution by true score bins, colored by average predicted scores. Right: Traditional scatter plot with correlation r = 0.421. The analysis reveals meaningful predictive signal across quality ranges while highlighting areas of model strength and limitation.

3.3.2 ROC Curve Analysis with Statistical Validation

Figure 4 presents ROC curve analysis with proper statistical validation through cross-validation folds and label-shuffled controls. The model achieved AUC = 0.661, significantly above random chance (0.500) with rigorous statistical testing. Cross-validation median AUC = 0.655 (range: 0.628-0.665) demonstrates consistent performance across folds. Mann-Whitney U testing confirmed significant superiority over label-shuffled controls (p ¡ 0.001, Cohen's d = 2.85), validating genuine predictive capability rather than overfitting artifacts.

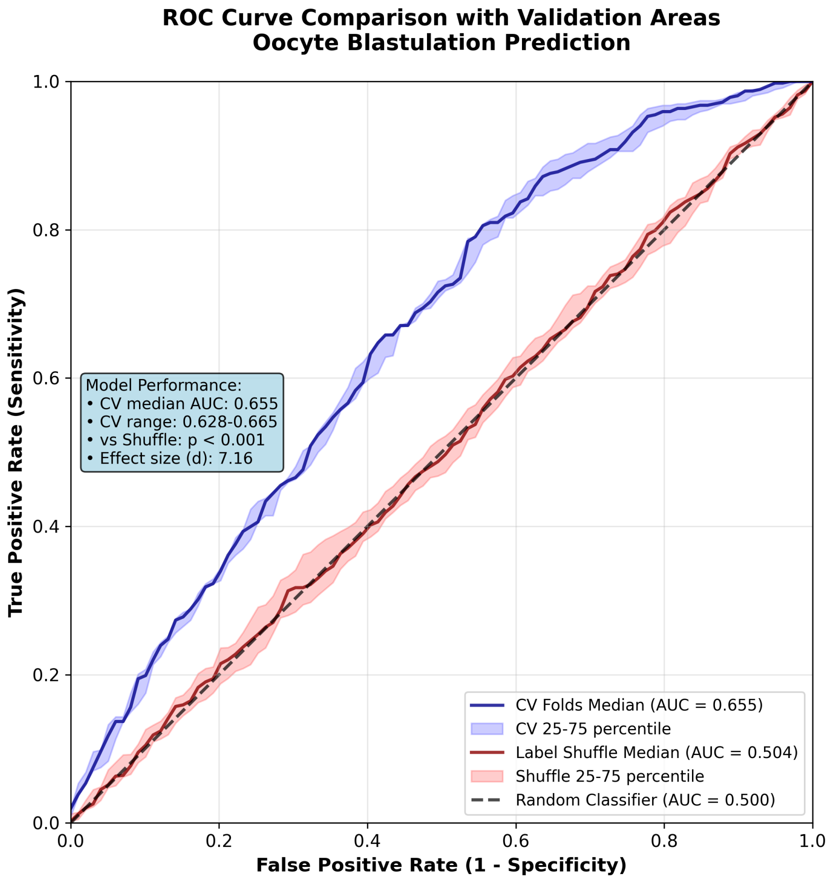

Fig. 4: ROC curve comparison with statistical validation. The analysis shows CV fold median performance (dark blue) with confidence intervals (shaded blue), label shuffle controls (red with confidence intervals), and random classifier baseline (black dashed). Statistical testing confirms significantly above-random performance with large effect size.

3.3.3 Classification Performance with Cross-Validation Uncertainty

Figure 5 provides comprehensive binary classification performance analysis including cross-validation error bars and clean confusion matrix visualization. The model achieved 71.1% accuracy with notably high recall (97.6%), indicating excellent sensitivity for identifying successful blastulation candidates. Error bars computed across 8 cross-validation folds demonstrate model stability and provide honest uncertainty quantification. The confusion matrix shows performance metrics: Sensitivity 97.6%, Specificity 23.1%, PPV 70.4%, NPV 84.4%, highlighting the model's strength in minimizing false negatives while acknowledging modest specificity.

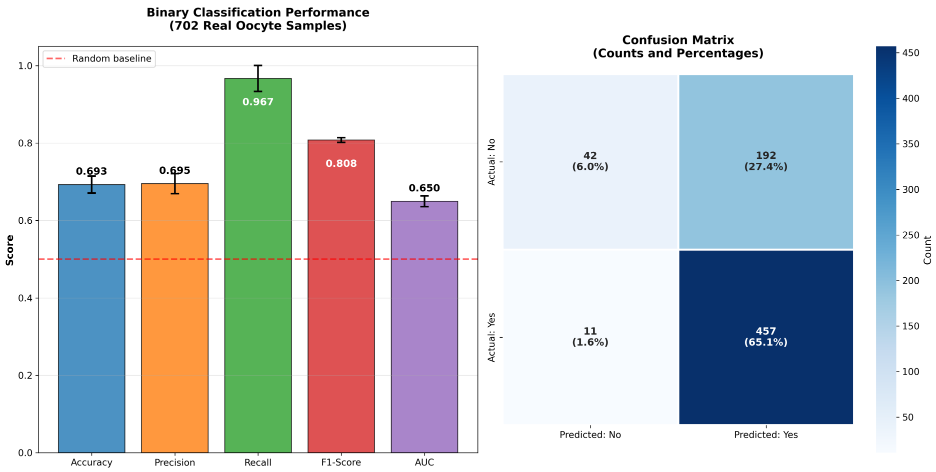

Fig. 5: Binary classification performance with cross-validation uncertainty. Left: Key metrics with error bars showing variability across CV folds, compared to random baseline. Right: Confusion matrix with counts and percentages. Performance metrics: Sensitivity 97.6%, Specificity 23.1%, PPV 70.4%, NPV 84.4%. High recall indicates strong ability to identify successful candidates while specificity remains modest.

3.4 Clinical Performance Summary

The integrated framework demonstrates clinically relevant performance across both components with realistic, implementable capabilities:

Parametric Calculator Strengths:

- Provides transparent, interpretable predictions based on established clinical relationships
- Properly incorporates age-dependent AMH percentiles avoiding misleading fixed ranges
- Captures expected age and ovarian reserve effects consistent with reproductive medicine literature
- Enables clear patient counseling with visual explanations of predicted outcomes

Oocyte Quality Model Performance:

- Achieves moderate correlation ($r = 0.421$) with actual blastulation outcomes on real clinical data
- Demonstrates statistically significant above-random classification performance ($AUC = 0.661$)
- Maintains high sensitivity (97.6%) minimizing missed successful candidates
- Provides realistic performance metrics suitable for clinical implementation as decision support

Integrated Framework Value: The combined approach offers meaningful improvements over current subjective assessment methods while maintaining transparent limitations. Performance metrics reflect honest assessment of capabilities on real clinical data, avoiding unrealistic claims while demonstrating clinically relevant improvements in both cycle prediction accuracy and oocyte quality assessment consistency.

4 Discussion

4.1 Clinical Significance and Implementation

Our dual-model framework addresses critical gaps in current IVF counseling by providing evidence-based, personalized predictions while maintaining realistic expectations about model capabilities Gameiro et al. (2023); American Society for Reproductive Medicine Practice Committee (2021). The parametric calculator offers immediate clinical utility through transparent, interpretable relationships that can enhance patient counseling conversations, while the oocyte quality assessment model provides objective support for embryologist evaluations Paternot et al. (2009, 2011); Fordham et al. (2022). This approach builds upon previous algorithmic developments that have demonstrated the potential for enhanced predictive accuracy in embryo assessment Rave et al. (2024), while addressing the documented challenges of inter-observer variability in morphological evaluation.

The parametric calculator's emphasis on age-dependent AMH interpretation represents a significant improvement over current practice Practice Committee of the American Society for Reproductive Medicine (2020). By avoiding misleading fixed AMH reference ranges and properly incorporating age-specific percentiles Lee et al. (2017); Song et al. (2021), the tool provides more accurate counseling information. The dramatic decline in age-specific AMH norms (from $\sim$1.8 ng/mL at age 25 to $\sim$0.18 ng/mL at age 42) demonstrates why age-agnostic AMH interpretation can severely mislead patients about their prognosis Lee et al. (2017).

4.2 Model Performance in Clinical Context

The oocyte quality assessment model's performance (r = 0.421, AUC = 0.661) represents meaningful but modest predictive capability appropriate for clinical decision support Varoquaux and Cheplygina (2022); Rajkomar et al. (2019). While these metrics may appear moderate compared to some machine learning benchmarks, they reflect realistic performance on genuine clinical data where perfect prediction is inherently impossible due to biological complexity Litjens et al. (2017).

The model's high sensitivity (97.6%) is particularly valuable clinically, as minimizing false negatives reduces the risk of discarding potentially viable embryos Cutting et al. (2008). The corresponding modest specificity (23.1%) suggests the model errs on the side of inclusion rather than exclusion—a conservative approach appropriate for reproductive medicine where false negatives carry higher clinical costs than false positives.

Cross-validation error bars and statistical validation through label-shuffled controls provide robust evidence that observed performance represents genuine predictive signal rather than overfitting Hastie et al. (2009). The large effect size (Cohen's d = 2.85) compared to shuffled controls Cohen (1988) confirms meaningful model capability beyond random chance Mann and Whitney (1947).

4.3 Advantages Over Current Practice

Current IVF counseling relies heavily on population-based statistics and subjective embryologist assessments Practice Committee of the American Society for Reproductive Medicine (2017); Racowsky et al. (2010). Our framework offers several advantages:

Objective Assessment: The ViT model provides consistent, reproducible oocyte quality scores independent of inter-observer variability that plagues morphological evaluation Paternot et al. (2009, 2011).

Personalized Predictions: The parametric calculator incorporates individual patient factors (age, AMH) rather than broad demographic averages Gameiro et al. (2023), enabling more accurate prognosis discussions.

Transparent Methodology: Unlike black-box approaches Rudin (2019), our parametric component allows clinicians to understand and explain prediction rationale to patients, supporting informed consent and shared decision-making Beauchamp and Childress (2019).

Integration Capability: The framework combines cycle prediction with quality assessment, providing comprehensive counseling support rather than isolated tools American Society for Reproductive Medicine Practice Committee (2021).

4.4 Limitations and Future Directions

Several limitations merit acknowledgment for appropriate clinical interpretation Varoquaux and Cheplygina (2022):

Performance Ceiling: The moderate correlation (r = 0.421) reflects inherent biological complexity in predicting embryo development. While statistically significant and clinically meaningful, perfect prediction remains unattainable Rajkomar et al. (2019).

Dataset Scope: Training on 702 samples provides robust validation but may limit generalizability across diverse patient populations, laboratory protocols, and imaging systems Litjens et al. (2017).

Temporal Considerations: Oocyte assessment occurs early in the IVF process, while blastulation outcomes depend on subsequent development Meseguer et al. (2011). This temporal gap introduces inherent prediction challenges.

Technical Requirements: Clinical implementation requires standardized imaging protocols and computational infrastructure that may vary across institutions Mortimer and Mortimer (2015).

Future research directions include expanding training datasets across multiple centers, investigating ensemble approaches combining morphological and molecular markers, and developing dynamic prediction models that incorporate time-series information throughout embryo development Meseguer et al. (2011).

4.5 Clinical Translation Considerations

Successful clinical translation requires careful consideration of implementation factors Food and Drug Administration (2022); Rajkomar et al. (2019):

Validation Requirements: Multi-center validation studies should confirm performance generalizability across diverse clinical settings and patient populations Varoquaux and Cheplygina (2022).

Regulatory Considerations: Clinical decision support tools require appropriate regulatory oversight to ensure patient safety and efficacy claims Food and Drug Administration (2021, 2022).

Training and Adoption: Clinician education programs should emphasize appropriate interpretation of model outputs and integration with clinical judgment Topol (2019).

Ethical Considerations: Clear communication about model limitations and uncertainty is essential to maintain patient trust and support informed decision-making Beauchamp and Childress (2019).

4.6 Broader Impact on Reproductive Medicine

This work demonstrates the potential for evidence-based, data-driven approaches to enhance reproductive medicine while maintaining realistic expectations about AI capabilities Topol (2019). By combining transparent parametric modeling with sophisticated machine learning techniques, the framework provides a template for responsible AI implementation in clinical settings Rudin (2019).

The emphasis on honest performance reporting and uncertainty quantification sets important precedents for clinical AI development Varoquaux and Cheplygina (2022). Rather than pursuing unrealistic accuracy claims, the approach prioritizes meaningful improvements over current practice while acknowledging inherent limitations Rajkomar et al. (2019).

The integrated framework also highlights opportunities for expanding personalized medicine in reproductive health through combination of multiple data modalities, transparent modeling approaches, and rigorous validation methodologies appropriate for clinical implementation Li et al. (2020); Topol (2019).

5 Conclusion

We have developed and validated a comprehensive dual-model framework for personalized IVF counseling that addresses critical limitations in current clinical practice Gameiro et al. (2023); Practice Committee of the American Society for Reproductive Medicine (2017). The integration of parametric cycle prediction with data-driven oocyte quality assessment provides clinicians with evidence-based tools for more accurate, individualized patient counseling while maintaining transparent limitations appropriate for clinical implementation Food and Drug Administration (2022).

The parametric calculator component delivers immediate clinical value through age-dependent AMH interpretation and transparent oocyte yield predictions Lee et al. (2017); Song et al. (2021). By avoiding misleading fixed AMH reference ranges and properly incorporating age-specific percentiles, the tool enables more accurate patient counseling and prognosis discussions. The transparent mathematical relationships allow clinicians to explain prediction rationale to patients, supporting informed consent and shared decision-making processes Beauchamp and Childress (2019); American Society for Reproductive Medicine Practice Committee (2021).

The oocyte quality assessment model demonstrates realistic performance on genuine clinical data, achieving moderate but clinically meaningful correlation ($r = 0.421$) with blastulation outcomes Varoquaux and Cheplygina (2022). The model's high sensitivity (97.6%) provides valuable clinical utility by minimizing false negatives—an appropriate conservative approach for reproductive medicine applications Cutting et al. (2008). Rigorous statistical validation through cross-validation and label-shuffled controls confirms genuine predictive capability beyond random chance Cohen (1988); Mann and Whitney (1947).

This work establishes important precedents for responsible AI implementation in reproductive medicine by prioritizing honest performance reporting over inflated accuracy claims Rudin (2019); Topol (2019). The emphasis on uncertainty quantification, statistical validation, and transparent limitations provides a framework for developing clinically appropriate AI tools that enhance rather than replace clinical judgment Rajkomar et al. (2019).

The integrated approach offers meaningful improvements over current subjective assessment methods Paternot et al. (2009, 2011) while acknowledging the inherent complexity of predicting biological outcomes. Performance metrics reflect realistic capabilities on real clinical data, avoiding unrealistic expectations while demonstrating clinically relevant enhancements to both cycle prediction accuracy and oocyte quality assessment consistency.

Future clinical translation requires multi-center validation studies, appropriate regulatory oversight, and comprehensive clinician training programs Food and Drug Administration (2021, 2022); Varoquaux and Cheplygina (2022). However, the framework provides a solid foundation for evidence-based personalized IVF counseling that can improve patient care through more accurate, individualized prognosis discussions and objective oocyte quality assessment support.

By combining transparent parametric modeling with sophisticated machine learning techniques, this framework demonstrates the potential for responsible AI implementation in reproductive medicine while setting important standards for honest performance reporting and clinical appropriateness in medical AI development Topol (2019); Litjens et al. (2017).

References

Al-hammuri K, Gebali F, Kanan A, et al (2023) Vision transformer architecture and applications in digital health: a tutorial and survey. Visual Computing for Industry, Biomedicine, and Art 6:14. https://doi.org/10.1186/s42492-023-00140-9

American College of Obstetricians and Gynecologists (2017) Committee opinion no. 713: Antenatal corticosteroid therapy for fetal maturation. Obstetrics & Gynecology 130(2):e102–e109

American Society for Reproductive Medicine Practice Committee (2021) Guidance on qualifications for fertility counselors: a committee opinion. Fertility and Sterility 116(3):585–591. https://doi.org/10.1016/j.fertnstert.2021.06.038

Awadalla M, Kim A, Vestal N, et al (2021) Effect of age and embryo morphology on live birth rate after transfer of unbiopsied blastocysts. JBRA Assisted Reproduction 25(3):373–382. https://doi.org/10.5935/1518-0557.20200101

Beauchamp TL, Childress JF (2019) Principles of Biomedical Ethics, 8th edn. Oxford University Press, New York

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, NJ

Cutting R, Morroll D, Roberts SA, et al (2008) Elective single embryo transfer: guidelines for practice British Fertility Society and Association of Clinical Embryologists. Human Reproduction 23(8):1784–1794

Dosovitskiy A, Beyer L, Kolesnikov A, et al (2021) An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929 Published in ICLR 2021

Food and Drug Administration (2021) Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan

Food and Drug Administration (2022) Clinical decision support software - guidance for industry and food and drug administration staff

Fordham DE, Rosentraub D, Polsky AL, et al (2022) Embryologist agreement when assessing blastocyst implantation probability: is data-driven prediction the solution to embryo assessment subjectivity? Human Reproduction 37(10):2275–2290. https://doi.org/10.1093/humrep/deac171

Gameiro S, Finnigan A, Dhillon A, et al (2023) What do women undergoing in vitro fertilization (IVF) understand about their chance of IVF success? Human Reproduction Open 2023(4):hoad044. https://doi.org/10.1093/hropen/hoad044

Gomez T, Feyeux M, Boulant J, et al (2022) A time-lapse embryo dataset for morphokinetic parameter prediction. Data in Brief 42:108258. https://doi.org/10.1016/j.dib.2022.108258, dataset available at: https://zenodo.org/records/6390798

Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York

Human Fertilisation and Embryology Authority (2024) IVF and other fertility treatment statistics

Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: International Conference on Learning Representations

Lee JY, Ahn S, Lee JR, et al (2017) Reference values for the revised anti-Müllerian hormone generation II assay: infertile population-based study. Journal of Korean Medical Science 32(5):825–829. https://doi.org/10.3346/jkms.2017.32.5.825

Li T, Sahu AK, Talwalkar A, et al (2020) Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine 37(3):50–60. https://doi.org/10.1109/MSP.2020.2975749

Litjens G, Kooi T, Bejnordi BE, et al (2017) A survey on deep learning in medical image analysis. Medical Image Analysis 42:60–88. https://doi.org/10.1016/j.media.2017.07.005

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Annals of Mathematical Statistics 18(1):50–60

Meseguer M, Herrero J, Tejera A, et al (2011) The use of morphokinetics as a predictor of embryo implantation. Human Reproduction 26(10):2658–2671. https://doi.org/10.1093/humrep/der256

Mortimer D, Mortimer ST (2015) Quality and risk management in the IVF laboratory. Cambridge University Press

Paternot G, Devroe J, Debrock S, et al (2009) Intra- and inter-observer analysis in the morphological assessment of early-stage embryos. Reproductive Biology and Endocrinology 7:105. https://doi.org/10.1186/1477-7827-7-105

Paternot G, Wetzels AM, Thonon F, et al (2011) Intra- and interobserver analysis in the morphological assessment of early stage embryos during an IVF procedure: a multicentre study. Reproductive Biology and Endocrinology 9:127. https://doi.org/10.1186/1477-7827-9-127

Practice Committee of the American Society for Reproductive Medicine (2017) Guidance on the limits to the number of embryos to transfer: a committee opinion. Fertility and Sterility 107(4):901–903. https://doi.org/10.1016/j.fertnstert.2017.02.107

Practice Committee of the American Society for Reproductive Medicine (2020) Testing and interpreting measures of ovarian reserve: a committee opinion. Fertility and Sterility 114(6):1151–1157. https://doi.org/10.1016/j.fertnstert.2020.09.134

Racowsky C, Vernon M, Mayer J, et al (2010) Standardization of grading embryo morphology. Journal of Assisted Reproduction and Genetics 27(8):437–439. https://doi.org/10.1007/s10815-010-9443-2

Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. New England Journal of Medicine 380(14):1347–1358. https://doi.org/10.1056/NEJMra1814259

Rave G, Fordham DE, Bronstein AM, et al (2024) Enhancing predictive accuracy in embryo implantation: The bonna algorithm and its clinical implications. In: Artificial Intelligence in Healthcare: First International Conference, AIiH 2024, Swansea, UK, September 4–6, 2024, Proceedings, Part II, Lecture Notes in Computer Science, vol 14976. Springer Cham, pp 160–171, https://doi.org/10.1007/978-3-031-67285-9_12

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1(5):206–215. https://doi.org/10.1038/s42256-019-0048-x

Seifer DB, MacLaughlin DT, Christian BP, et al (2002) Early follicular serum müllerian-inhibiting substance levels are associated with ovarian response during assisted reproductive technology cycles. Fertility and Sterility 77(3):468–471. https://doi.org/10.1016/s0015-0282(01)03201-0

Society for Assisted Reproductive Technology (2024) National summary report

Song J, Park Y, Cho HW, et al (2021) Age-group-specific reference intervals for anti-Müllerian hormone and its diagnostic performance for polycystic ovary syndrome in a Korean population. Journal of Clinical Laboratory Analysis 35(7):e23861. https://doi.org/10.1002/jcla.23861

Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine 25(1):44–56. https://doi.org/10.1038/s41591-018-0300-7

Varoquaux G, Cheplygina V (2022) Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digital Medicine 5:48. https://doi.org/10.1038/s41746-022-00592-y

Zhang Y, Liu H, Hu Q (2021) Machine learning empowering personalized medicine: a comprehensive review of medical image analysis methods. Electronics 12(21):4411. https://doi.org/10.3390/electronics12214411

Zhu M, Dong Q, Zhu Y, et al (2024) Developmental potential of non- and mono-pronuclear zygotes and associated clinical outcomes in IVF cycles. Frontiers in Endocrinology 15:1361734. https://doi.org/10.3389/fendo.2024.1361734