Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multimodal Deep Learning Integrating Day 3 Embryo Images and Clinical Data Predicts Pregnancy Outcomes After Fresh Single Embryo Transfer

Jialin Wu
Sixth Affiliated Hospital of Sun Yat-sen University
Yanfang Wang
Sixth Affiliated Hospital of Sun Yat-sen University
Tingting Li
Sixth Affiliated Hospital of Sun Yat-sen University
Xiaoyan Liang
Sixth Affiliated Hospital of Sun Yat-sen University
Rui Huang
Sixth Affiliated Hospital of Sun Yat-sen University
Linan Xu
xulinan@mail.sysu.edu.cn

Sixth Affiliated Hospital of Sun Yat-sen University

Article

Keywords: Single embryo transfer, clinical pregnancy, live birth, multimodal deep learning, in vitro fertilization

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10710514/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Multimodal Deep Learning Integrating Day 3 Embryo Images and Clinical Data Predicts Pregnancy Outcomes After Fresh Single Embryo Transfer

Jialin Wu M.D.[a,b], Yanfang Wang B.S.[a,b], Tingting Li M.D., Ph.D.[a,b], Xiaoyan Liang M.D., Ph.D. [a,b], Rui Huang M.D., Ph.D.[a,b*] and Linan Xu M.D., Ph.D.[a,b*]

[a] Reproductive Medicine Center, Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.

[b] Biomedical Innovation Center, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, China.

Corresponding Author:

Rui Huang

Reproductive medicine center, the Sixth Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-sen University

17, Shou Gou Ling Road, Guangzhou, China 510610.

Phone: 86-20-38048012

Email: huangr57@mail.sysu.edu.cn

Linan Xu

Reproductive medicine center, the Sixth Affiliated Hospital, Zhongshan School of Medicine, Sun Yat-sen University

17, Shou Gou Ling Road, Guangzhou, China 510610.

Phone: 86-20-38048010

Email: xulinan@mail.sysu.edu.cn

Rui Huang and Linan Xu are co-corresponding authors for this work.

Abstract

This retrospective cohort study evaluated whether a multimodal deep learning model integrating Day 3 (D3) embryo images and clinical parameters could improve predictions of clinical pregnancy (CP), live birth (LB), and miscarriage following fresh single embryo transfer (SET), compared to unimodal approaches. A total of 1,207 fresh D3 SET cycles performed between January 2024 and August 2025 at a single center were analyzed. We developed an ImprovedFusion model combining a convolutional neural network for image feature extraction and a multilayer perceptron for clinical data, alongside unimodal benchmarks (ClinicalMLP and ImageResNet). Model performance was assessed using area under the receiver operating characteristic curve (AUC), sensitivity, and specificity, with SHAP and Grad-CAM providing interpretability. The CP rate was 44.32% (535/1,207), the LB rate was 35.82% (423/1,181; 26 lost to follow-up), and pregnancy loss occurred in 25.17% (151/600) of biochemical pregnancies. ImprovedFusion significantly outperformed both unimodal models across all endpoints. For CP, the multimodal model achieved an AUC of 0.754 (95% CI: 0.683–0.826; sensitivity: 0.889, specificity: 0.436), versus

0.689 and 0.592 for unimodal models. For LB, the AUC was 0.721 (95% CI: 0.647–0.795; sensitivity: 0.750), versus 0.665 and 0.626. For miscarriage, the AUC was 0.702 (95% CI: 0.576–0.828), versus 0.679 and 0.553. SHAP identified endometrial thickness, maternal age, trigger-day hormones, and image features as top predictors, while Grad-CAM highlighted greater attention to blastomere regions over fragmentation, suggesting cell number predicts better. Following external validation, this framework may offer a cost-effective, interpretable tool to enhance embryo selection and reproductive outcomes.

Keywords

Single embryo transfer; clinical pregnancy; live birth; multimodal deep learning; in vitro fertilization

Introduction

Although in vitro fertilization (IVF) has become a cornerstone in the treatment of infertility, the live birth (LB) rate per single cycle remains below 30% [1], indicating substantial scope for clinical optimization. Embryo selection is a pivotal determinant of IVF success. Traditionally, this process has relied heavily on the morphological assessment of embryos by embryologists, an approach that is inherently subjective and subject to considerable inter-observer variability [2], thereby limiting its predictive accuracy [3].

In recent years, the advent of deep learning, particularly convolutional neural networks (CNNs), has introduced novel and objective methodologies for embryo evaluation [4]. Nevertheless, existing research has predominantly concentrated on blastocyst-stage assessment on day 5/6 [5–7], with comparatively limited attention directed toward cleavage-stage embryos on Day 3 (D3). In clinical practice, D3 embryo transfer is particularly meaningful for patients with reduced blastocyst formation potential (e.g., advanced maternal age) or those with a limited number of embryos [8]. Currently, the selection of D3 embryos relies primarily on morphological grading, but this method has notable limitations in predicting subsequent embryonic developmental potential. To compensate for this uncertainty, clinicians often tend to transfer two or even more D3 embryos to maximize the chance of at least one successful implantation. However, this practice significantly increases the risk of multiple pregnancies and associated maternal and fetal complications [9]. Therefore, there is an urgent need to develop more accurate indicators of embryonic developmental potential to guide elective single embryo transfer (SET), thereby effectively reducing the risk of multiple pregnancies while maintaining optimal clinical outcomes.

Embryo implantation potential is affected by a complex interplay of multiple factors, including maternal age, endocrine status, immune profile, and endometrial

receptivity [10]. Traditional statistical models are often inadequate for capturing the nonlinear interactions between clinical variables and embryonic morphological features [11]. Moreover, existing artificial intelligence models frequently fail to integrate clinical parameters comprehensively with image-based data [12]. To address these limitations, the present study developed a multimodal CNN model that incorporates both embryonic morphological images and corresponding clinical data, with the objective of enhancing the predictive performance for D3 embryo transfer outcomes. The specific aims of this study are: (1) to compare the predictive efficacy of models utilizing different data modalities; (2) to validate the advantages of the multimodal fusion model over traditional approaches; (3) to identify key clinical and morphological variables influencing pregnancy outcomes using SHAP analysis;and (4) to provide a clinically interpretable tool for individualized embryo selection, with potential benefits for patients undergoing D3 embryo transfer.

Materials and Methods

1. Study Population and Dataset

This retrospective study analyzed data from 1,207 fresh cleavage-stage SET cycles performed on D3 at the Reproductive Medicine Center of the Sixth Affiliated Hospital of Sun Yat-sen University between January 2024 and August 2025. Inclusion criteria were: (1) clear static D3 embryo images; (2) complete medical records; and (3) documented outcomes including clinical pregnancy (CP), LB, and miscarriage. All 1,207 cycles were included for CP analysis; miscarriage analysis was restricted to the 600 cycles achieving biochemical pregnancy; and LB analysis included 1,181 cycles after excluding 26 lost to follow-up.

The primary outcomes were CP and LB. CP was defined as the presence of an intrauterine gestational sac confirmed by ultrasound. LB was defined as delivery of a live infant beyond 28 weeks of gestation. The secondary outcome was miscarriage, which included biochemical miscarriage (loss of a biochemical pregnancy before ultrasound confirmation) and clinical miscarriage (loss of a clinical pregnancy before 20 weeks of gestation).

A total of 28 clinical features were collected, covering: (1) female factors, including age, body mass index (BMI), infertility type and duration, obstetric history (gravidity, parity, abortion, and ectopic pregnancy), and basal serum hormone levels (follicle-stimulating hormone [FSH], estradiol [E_2], prolactin, luteinizing hormone [LH], and testosterone [T]); (2) male factors, including age and parameters from the initial semen analysis (volume, pH, proportion of progressively motile sperm, sperm density, and normal morphology rate); and (3) cycle characteristics, including the

ovarian stimulation protocol, serum hormone levels on the trigger day (E_2, LH, progesterone), the number of oocytes retrieved, the number of two pronuclei (2PN) fertilized oocytes, the number of high-quality embryos (HQE), endometrial thickness (EMT), and endometrial pattern on the trigger day.

2. Dataset Partition

The dataset was split using stratified sampling into training, validation, and test sets in a ratio of 7:1.5:1.5, preserving the proportion of positive outcomes across all splits (Supplemental Table 1).

3. Clinical Feature Screening

Features with univariate association ($p \leq 0.05$) to the outcome were retained. Continuous variables were compared using t-test or Mann-Whitney U test; categorical variables using χ^2 or Fisher's exact test. Analyses were performed using R 4.2.1 and Python 3.9.

4. Model Development

Three models were developed for each prediction task (CP, LB, and miscarriage):

Clinical-only model (ClinicalMLP): A three-layer fully connected network with batch normalization, ReLU activation, and dropout regularization (rate 0.3) after each hidden layer.

Image-only model (ImageResNet): A ResNet-50 architecture pretrained on ImageNet, with the final layers fine-tuned on our dataset. The classification head consisted of two fully connected layers (2,048→512→2) with dropout.

Multimodal fusion model (ImprovedFusion): This model integrated image and clinical features using an adaptive weighting mechanism that learned the relative importance of each modality. The image branch was a ResNet-50 pretrained on ImageNet; the clinical branch was a two-layer network. Their outputs were weighted by learned attention scores, concatenated, and passed to a three-layer classifier.

5. Training and Optimization

All models were implemented in PyTorch 2.0 and trained on NVIDIA GPUs. Weighted cross-entropy loss was used to address class imbalance. The AdamW optimizer was applied with a learning rate of 5×10^{-4} for the clinical-only model and 1×10^{-4} for the image-only and multimodal models. All models were trained for up to 60 epochs with early stopping (patience = 10) based on validation area under the curve (AUC).

6. Threshold Optimization

Optimal classification thresholds were determined on the validation set by searching from 0.1 to 0.9 in steps of 0.05 to maximize the F1-score. The selected thresholds were then applied to the test set.

7. Model Performance Evaluation

Primary metrics were AUC, sensitivity, and specificity. Secondary metrics included accuracy, precision, F1-score, and balanced accuracy. AUC differences were compared using DeLong's test (two-tailed $p < 0.05$). SHAP values were computed for model interpretability.

8. Model Interpretability with Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize image regions influencing model predictions. Heatmaps were generated for both ImageResNet and ImprovedFusion and overlaid on original images. Prediction confidence scores were compared using the Wilcoxon signed-rank test. Visualizations were generated using the PyTorch Grad-CAM library.

9. Ethical Approval

This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Sixth Affiliated Hospital of Sun Yat-sen University (Approval Number: 2025ZSLYEC-721). The requirement for informed consent was waived due to the retrospective nature of the study and the use of fully anonymized data, with no additional interventions involving human participants.

Results

1. Study Population and Dataset Characteristics

A total of 1,207 fresh SET cycles were included in this study. Demographic and clinical information is summarized in Table 1 (by CP outcome), Supplemental Table 2 (by LB outcome), and Supplemental Table 3 (by miscarriage outcome). Figure 1 provides a flowchart detailing the progression of cycles from the initial cohort to each pregnancy outcome, with corresponding sample sizes. Clinical features were preselected using univariate analysis and varied by prediction task. A detailed summary of the features included for each outcome is provided in Supplemental Table 4.

2. Model Performance on Clinical Pregnancy, Live Birth, and Miscarriage Prediction

On the independent test set, ImprovedFusion consistently outperformed single-modal baselines across all three outcomes (Table 2). For CP prediction,

ImprovedFusion achieved the highest AUC (0.754, 95% CI: 0.683–0.826), sensitivity (0.889), and balanced accuracy (0.662); DeLong's test showed significant improvement over ClinicalMLP ($p = 0.050$) and ImageResNet ($p < 0.05$). For LB prediction, ImprovedFusion attained an AUC of 0.721 (95% CI: 0.647–0.795; ClinicalMLP: 0.665 [95% CI: 0.586–0.745], $p = 0.119$; ImageResNet: 0.626 [95% CI: 0.540–0.711], $p = 0.058$), with superior balanced accuracy (0.682) and F1-score (0.615). For miscarriage prediction, ImprovedFusion achieved an AUC of 0.702 (95% CI: 0.576–0.828; ClinicalMLP: 0.679 [95% CI: 0.551–0.808], $p = 0.614$; ImageResNet: 0.553 [95% CI: 0.404–0.703], $p = 0.180$), outperforming both single-modal baselines.

3. Multimodal Fusion Gain

Single-modal models exhibited distinct diagnostic biases (Table 2 & Supplemental Figure 1), whereas ImprovedFusion achieved superior balance across sensitivity and specificity. For CP prediction, the fusion model improved AUC by 9.4% over ClinicalMLP and 27.4% over ImageResNet. For LB prediction, AUC improved by 8.4% over ClinicalMLP and 15.2% over ImageResNet. For miscarriage prediction, AUC improved by 3.4% over ClinicalMLP and 26.9% over ImageResNet. Notably, ImprovedFusion correctly identified 88.9% of CPs and 75.0% of LBs, whereas ImageResNet exhibited a high false-positive rate of 77.2% for CP and a low sensitivity of 0.719 for LB.

4. Feature Importance Ranking via SHAP Analysis

SHAP analysis was employed to quantify feature contributions to model predictions (Supplemental Figure 2). For CP prediction, the top five influential features were E_2 levels on trigger day (0.109), LH levels on trigger day (0.101), the number of oocytes retrieved (0.099), semen volume (0.094), and HQE (0.082). For LB prediction, EMT ranked first (0.133), followed by female age (0.131), LH levels on trigger day (0.115), E_2 levels on trigger day (0.103), and image features (0.095). For miscarriage prediction, the leading contributors were basal LH levels, female age, EMT, semen volume, and image features.

5. Visualization of Model Attention via Grad-CAM

To investigate the regions of interest prioritized by the CNN and assess whether the incorporation of clinical data modulates attentional focus, we employed Grad-CAM. Figure 2 presents comparative visualizations of representative D3 cleavage-stage embryos showing: (a) the original image, (b) the heatmap from ImageResNet, and (c) the heatmap from ImprovedFusion.

Quantitative analysis revealed that ImprovedFusion significantly outperformed the

ImageReset model in prediction confidence for all tasks, achieving higher confidence scores for CP (0.979 vs. 0.508, $p < 0.001$), LB (0.798 vs. 0.709, $p < 0.001$) and miscarriage prediction (0.500 vs. 0.422, $p < 0.001$).

Qualitatively, ImageResNet exhibited diffuse and scattered activation patterns, with high-intensity regions distributed around the embryo. In contrast, ImprovedFusion demonstrated more concentrated to areas of cell–cell contact and overlapping blastomeres. This shift suggests that the integration of clinical information might guide the model to attend to morphologically meaningful structures associated with developmental potential.

Discussion

The accurate prediction of embryo implantation potential remains a cornerstone challenge in reproductive medicine. In this study, we developed and validated a multimodal deep learning model that integrates cleavage-stage embryo images and clinical data to predict pregnancy outcomes after fresh SET on D3. The proposed ImprovedFusion model outperformed the unimodal baseline models in all three prediction tasks, confirming the complementary value of embryo morphological information and clinical data.

In predicting CP and LB, the ClinicalMLP model achieved AUCs ranging from 0.665 to 0.689, suggesting that clinical parameters alone cannot fully capture early implantation events [11,13]. While CNNs have been widely applied to blastocyst image assessment [3,12,14–18], their LB prediction AUC using only blastocyst images is approximately 0.65 (12,18). Our study obtained similar results using only cleavage-stage images (AUC = 0.626), consistent with these prior reports. However, studies specifically targeting D3 cleavage-stage embryos are scarce, and the association between their morphological features and pregnancy outcomes remains unclear. Neither clinical parameters nor static embryo images can reflect the dynamic embryo–endometrium interaction. Many non-pregnant embryos may retain viability but fail to implant due to factors such as endometrial receptivity [19]. The poorer performance of ImageResNet confirms the limited value of single-modality approaches using static images. In contrast, the ImprovedFusion model raised the AUC to 0.702–0.754 across the three prediction tasks, representing a relative improvement of 3.4-9.4% over the clinical-only model. Notably, the ClinicalMLP models' AUCs of approximately 0.6 for all three outcomes, indicating that traditional clinical indicators poorly assess post-implantation developmental risk. This risk likely depends on dynamic embryo–endometrial crosstalk, with some predictive information embedded in morphological features [20,21]. By incorporating image

data, the sensitivity of CP ImprovedFusion model was enhanced to 0.889 (relative improvement 10.8%), and its accuracy (0.662) exceeded that of several previous multimodal studies (0.643-0.652) [3,16]. For instance, the logistic model for miscarriage prediction developed by Ouyang et al. achieved a sensitivity of only 62.8% [22], whereas our ImprovedFusion achieved 65.1%, suggesting a potential advantage in identifying high-risk cycles. Together, these results suggest that conventional logistic regression models, still widely used in reproductive centers, may underestimate outcome predictability.

SHAP analysis revealed that only 7, 6, and 4 core features were needed to predict CP, LB, and miscarriage, respectively, capturing over 60% of the predictive information. This indicates substantial redundancy among the many indicators included in current conventional models. A simplified multimodal model is not only feasible but also reduces data collection and quality control costs, facilitating deployment to resource-limited reproductive centers.

Among the common predictive factors, serum E_2 and LH levels on trigger day were critical for both CP and LB, suggesting that pre-transfer endocrine status influences pregnancy establishment and maintenance. Elevated E_2 and LH levels are associated with higher oocyte yield and potentially more HQE [23–25]. However, excessively high E_2 levels impair endometrial receptivity and oocyte quality [26], while low LH levels may adversely affect oocyte development and endometrial growth [27]. Thus, hormonal levels exhibit a complex, non-linear relationship with outcomes. Among clinical features, HQE ranked fifth for CP, aligning with clinical experience that the number of HQE reflects both oocyte yield and developmental competence. Nevertheless, this metric alone does not capture individual embryo viability. EMT emerged as the top contributor for LB prediction. Previous studies have suggested that a thin endometrium may expose the embryo to higher vascular and oxygen concentrations in the basal endometrium, generating reactive oxygen species that create a suboptimal uterine environment, thereby impeding implantation and placentation and ultimately leading to impaired fetal growth [28,29]. More importantly, image features ranked fifth for LB and eighth for CP, suggesting that morphological information contributes more substantially to LB prediction than to CP alone. This indicates that static embryo images inform not only implantation potential but also the subsequent ability to sustain a CP to LB [12].

Time-lapse photography, a popular embryo selection tool, has been reported to achieve CP prediction AUCs of 0.68–0.71 [3,16,17] and LB prediction AUCs of approximately 0.65–0.66 [12,18]. Notably, our ImprovedFusion model surpassed these ranges, achieving 0.754 for CP, 0.721 for LB and 0.702 for miscarriage.

Redundant frames in time-lapse videos may introduce noise, leading to overfitting and limiting predictive accuracy [30–32]. Moreover, the high cost of dedicated time-lapse equipment restricts its widespread clinical adoption. In contrast, our framework uses static images obtained from routine morphological assessment, offering a more accessible and cost-effective alternative.

Another key advantage of multimodal fusion is its ability to overcome the inherent sensitivity–specificity trade-off of single-modal models. In CP prediction, the ImprovedFusion model maintained a high sensitivity of 0.889 while achieving a specificity of 0.436; in LB prediction, it also exhibited balanced performance. This trade-off is particularly valuable for clinical screening: high sensitivity helps avoid missing high-risk cycles, whereas moderate specificity reduces unnecessary interventions. A key finding from Grad-CAM visualization is that the inclusion of clinical data shifted the model's attention to focus more on blastomere regions. This suggests that clinical characteristics guide the model to prioritize biologically meaningful morphological structures. Although Stylianou et al. highlighted the degree of fragmentation, rather than cell number, as a strong predictor of pregnancy following D2/3 embryo transfer, our Grad-CAM analysis reveals, in contrast, that the model focuses more on blastomeres regions representing cell number when differentiating between cycles resulting in CP and LB [33]. This attention bias may be explained by the "pseudo-fragmentation" phenomenon – some fragments transiently appear during cell division and subsequently disappear [34]. Cell number has been consistently shown to correlate positively with implantation and LB rates for D3 transfers, particularly for 8-cell embryos [35,36]. Furthermore, a low blastomere count (≤5) is a known predictor of early pregnancy failure [37,38]. By integrating clinical parameters, the model gains a favorable background framework that helps ignore maternal-related noise (e.g., maternal age, EMT) and focus on intrinsic embryonic developmental potential. Notably, blastomere composition and arrangement are also closely linked to chromosomal ploidy status [37].

Our multimodal framework demonstrated consistent superiority across three clinically relevant endpoints. This cross-outcome generalization suggests that the model captures core embryo quality characteristics rather than outcome-specific artifacts. Performance differences among outcomes may reflect distinct biological determinants: CP is more directly predicted by D3 morphology, whereas LB and miscarriage are additionally influenced by later developmental events and maternal factors. Future studies could develop outcome-specific architectures to further improve accuracy. Importantly, the ImprovedFusion model markedly reduced attention to fragmented regions, indicating that multimodal learning effectively

suppresses low-predictive-value morphological noise, thereby constructing safer and more robust representations for clinical embryo assessment.

This study has several limitations. First, as a single-center retrospective analysis, the findings may be subject to selection bias, and generalizability requires validation in multicenter, prospective trials. Second, the Grad-CAM technique provides relatively coarse localization maps; future work should incorporate more advanced interpretability methods to precisely identify subcellular feature. Third, the miscarriage prediction analysis was constrained by a limited number of positive event, which likely contributed to the suboptimal specificity and precision observed. Thus, these results should be considered preliminary, and larger, more balanced cohorts are needed for validation.

Conclusions

This study presents a multimodal deep learning model that integrates D3 embryo images with clinical data to predict SET outcomes. The ImprovedFusion model consistently outperformed single-modal baselines in pregnancy outcome prediction, achieving improved sensitivity–specificity balance. SHAP analysis identified trigger-day E_2 and LH level, and image features as core predictors, supporting streamlined clinical assessment. Grad-CAM visualization revealed that the inclusion of clinical data refocused model attention onto blastomere regions, underscoring the central role of cell number in developmental potential and providing computational evidence that cell count outweighs fragmentation in predicting pregnancy outcomes. Following multicenter prospective validation, this model offers a promising data-driven tool for individualized embryo selection.

Acknowledgements

The authors are grateful to the participants of this study who made this work possible.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Author contributions

J.L.W. was involved in the conception and design of the study, methodology, software development, formal analysis, interpretation of the data and drafting of the manuscript. Y.F.W. was involved in the acquisition and data curation of the data. T.T.L. was involved in the investigation and validation for the study. X.Y.L. was involved in

the provision of resources and supervision. R.H. was involved in the interpretation of the data, project administration, and critically revised the manuscript for important intellectual content, and provided final approval for publication. L.N.X. was involved in the conception and design of the study, supervision, and critically revised the manuscript for important intellectual content, and provided final approval for publication.

Funding

No external funding was used for this study.

Declarations

Competing interests

The authors declare no competing interests.

Reference

1. European IVF Monitoring Consortium (EIM), for the European Society of Human Reproduction and Embryology (ESHRE) *et al.* ART in Europe, 2018: results generated from European registries by ESHRE. *Hum Reprod Open* **2022**, hoac022 (2022).
2. Baxter Bendus, A. E., Mayer, J. F., Shipley, S. K. & Catherino, W. H. Interobserver and intraobserver variation in day 3 embryo grading. *Fertil Steril* **86**, 1608–1615 (2006).
3. VerMilyea, M. *et al.* Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF. *Hum Reprod* **35**, 770–784 (2020).
4. Gilboa, D. *et al.* Application of a methodological framework for the development and multicenter validation of reliable artificial intelligence in embryo evaluation. *Reprod Biol Endocrinol* **23**, 16 (2025).
5. Tran, D., Cooke, S., Illingworth, P. J. & Gardner, D. K. Deep learning as a predictive

tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod* **34**, 1011–1018 (2019).

6. Chavez-Badiola, A. *et al.* Use of artificial intelligence embryo selection based on static images to predict first-trimester pregnancy loss. *Reproductive BioMedicine Online* **49**, (2024).
7. Salih, M. *et al.* Deep learning classification integrating embryo images with associated clinical information from ART cycles. *Sci Rep* **15**, 17585 (2025).
8. Bergh, C., Kamath, M. S., Wang, R. & Lensen, S. Strategies to reduce multiple pregnancies during medically assisted reproduction. *Fertil Steril* **114**, 673–679 (2020).
9. Katz, J. *et al.* Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. *Lancet* **382**, 417–425 (2013).
10. Medenica, S. *et al.* The Future Is Coming: Artificial Intelligence in the Treatment of Infertility Could Improve Assisted Reproduction Outcomes-The Value of Regulatory Frameworks. *Diagnostics (Basel)* **12**, 2979 (2022).
11. Gao, H., Liu, D.-E., Li, Y., Wu, X. & Tan, H. Early prediction of live birth for assisted reproductive technology patients: a convenient and practical prediction model. *Sci Rep* **11**, 331 (2021).
12. Nagaya, M. & Ukita, N. Embryo Grading With Unreliable Labels Due to Chromosome Abnormalities by Regularized PU Learning With Ranking. *IEEE Trans Med Imaging* **41**, 320–331 (2022).

13. Zhu, H., Zhao, C., Xiao, P. & Zhang, S. Predicting the Likelihood of Live Birth in Assisted Reproductive Technology According to the Number of Oocytes Retrieved and Female Age Using a Generalized Additive Model: A Retrospective Cohort Analysis of 17,948 Cycles. *Front Endocrinol (Lausanne)* **12**, 606231 (2021).
14. Kragh, M. F., Rimestad, J., Lassen, J. T., Berntsen, J. & Karstoft, H. Predicting Embryo Viability Based on Self-Supervised Alignment of Time-Lapse Videos. *IEEE Trans Med Imaging* **41**, 465–475 (2022).
15. Berntsen, J., Rimestad, J., Lassen, J. T., Tran, D. & Kragh, M. F. Robust and generalizable embryo selection based on artificial intelligence and time-lapse image sequences. *PLoS One* **17**, e0262661 (2022).
16. Enatsu, N. *et al.* A novel system based on artificial intelligence for predicting blastocyst viability and visualizing the explanation. *Reprod Med Biol* **21**, e12443 (2022).
17. Loewke, K. *et al.* Characterization of an artificial intelligence model for ranking static images of blastocyst stage embryos. *Fertility and Sterility* **117**, 528–535 (2022).
18. Miyagi, Y., Habara, T., Hirata, R. & Hayashi, N. Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image. *Reprod Med Biol* **18**, 204–211 (2019).
19. Makrigiannakis, A., Makrygiannakis, F. & Vrekoussis, T. Approaches to Improve Endometrial Receptivity in Case of Repeated Implantation Failures. *Front Cell Dev Biol* **9**, 613277 (2021).
20. Ashary, N., Tiwari, A. & Modi, D. Embryo Implantation: War in Times of Love.

Endocrinology **159**, 1188–1198 (2018).

21. Mahutte, N. *et al.* Optimal endometrial thickness in fresh and frozen-thaw in vitro fertilization cycles: an analysis of live birth rates from 96,000 autologous embryo transfers. *Fertil Steril* **117**, 792–800 (2022).
22. Ouyang, Y., Peng, Y., Zhang, S., Gong, F. & Li, X. A simple scoring system for the prediction of early pregnancy loss developed by following 13,977 infertile patients after in vitro fertilization. *Eur J Med Res* **28**, 237 (2023).
23. Kosmas, I. P., Kolibianakis, E. M. & Devroey, P. Association of estradiol levels on the day of hCG administration and pregnancy achievement in IVF: a systematic review. *Hum Reprod* **19**, 2446–2453 (2004).
24. Sunkara, S. K. *et al.* Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. *Hum Reprod* **26**, 1768–1774 (2011).
25. Malchau, S. S. *et al.* Cumulative live birth rate prognosis based on the number of aspirated oocytes in previous ART cycles. *Hum Reprod* **34**, 171–180 (2019).
26. Karatasiou, G. I. *et al.* Is the probability of pregnancy after ovarian stimulation for IVF associated with serum estradiol levels on the day of triggering final oocyte maturation with hCG? A systematic review and meta-analysis. *J Assist Reprod Genet* **37**, 1531–1541 (2020).
27. Luo, X. *et al.* Low Endogenous LH on the COS Initiation Day of a GnRH-Agonist Regimen Increases the Risk of Early Pregnancy Loss and Adverse ART Outcomes. *Front Endocrinol (Lausanne)* **13**, 830567 (2022).

28. Xu, J. *et al.* The Effects of Endometrial Thickness on Pregnancy Outcomes of Fresh IVF/ICSI Embryo Transfer Cycles: An Analysis of Over 40,000 Cycles Among Five Reproductive Centers in China. *Front Endocrinol (Lausanne)* **12**, 788706 (2021).
29. Schmiech, K., Li, M., Chen, L. X., Dow, M. P. & Baker, V. L. Association of endometrial thickness with live birth rate: a study using the Society for Assisted Reproductive Technology Clinic Outcome Reporting System. *Fertil Steril* **124**, 79–87 (2025).
30. Zhu, C. *et al.* Fine-Grained Video Categorization with Redundancy Reduction Attention. in *Computer Vision – ECCV 2018* (eds Ferrari, V., Hebert, M., Sminchisescu, C. & Weiss, Y.) vol. 11209 139–155 (Springer International Publishing, Cham, 2018).
31. Wu, X., Yang, F., Zhou, T. & Lin, X. Rethinking the Impacts of Overfitting and Feature Quality on Small-scale Video Classification. in *Proceedings of the 29th ACM International Conference on Multimedia* 4760–4764 (ACM, Virtual Event China, 2021). doi:10.1145/3474085.3479226.
32. Tao, H., Lu, M., Hu, Z., Xin, Z. & Wang, J. Attention-Aggregated Attribute-Aware Network With Redundancy Reduction Convolution for Video-Based Industrial Smoke Emission Recognition. *IEEE Trans. Ind. Inf.* **18**, 7653–7664 (2022).
33. Stylianou, C., Critchlow, D., Brison, D. R. & Roberts, S. A. Embryo morphology as a predictor of IVF success: An evaluation of the proposed UK ACE grading scheme for cleavage stage embryos. *Human Fertility* **15**, 11–17 (2012).
34. Van Blerkom, J., Davis, P. & Alexander, S. A microscopic and biochemical study of fragmentation phenotypes in stage-appropriate human embryos. *Hum Reprod* **16**,

719–729 (2001).

35. Wu, J., Zhang, J., Kuang, Y., Chen, Q. & Wang, Y. The effect of Day 3 cell number on pregnancy outcomes in vitrified-thawed single blastocyst transfer cycles. *Hum Reprod* **35**, 2478–2487 (2020).

36. Racowsky, C. *et al.* National collection of embryo morphology data into Society for Assisted Reproductive Technology Clinic Outcomes Reporting System: associations among day 3 cell number, fragmentation and blastomere asymmetry, and live birth rate. *Fertil Steril* **95**, 1985–1989 (2011).

37. Prados, F. J., Debrock, S., Lemmen, J. G. & Agerholm, I. The cleavage stage embryo. *Human Reproduction* **27**, i50–i71 (2012).

38. Working Group on the update of the ESHRE/ALPHA Istanbul Consensus *et al.* The Istanbul consensus update: a revised ESHRE/ALPHA consensus on oocyte and embryo static and dynamic morphological assessment†,‡. *Hum Reprod* **40**, 989–1035 (2025).

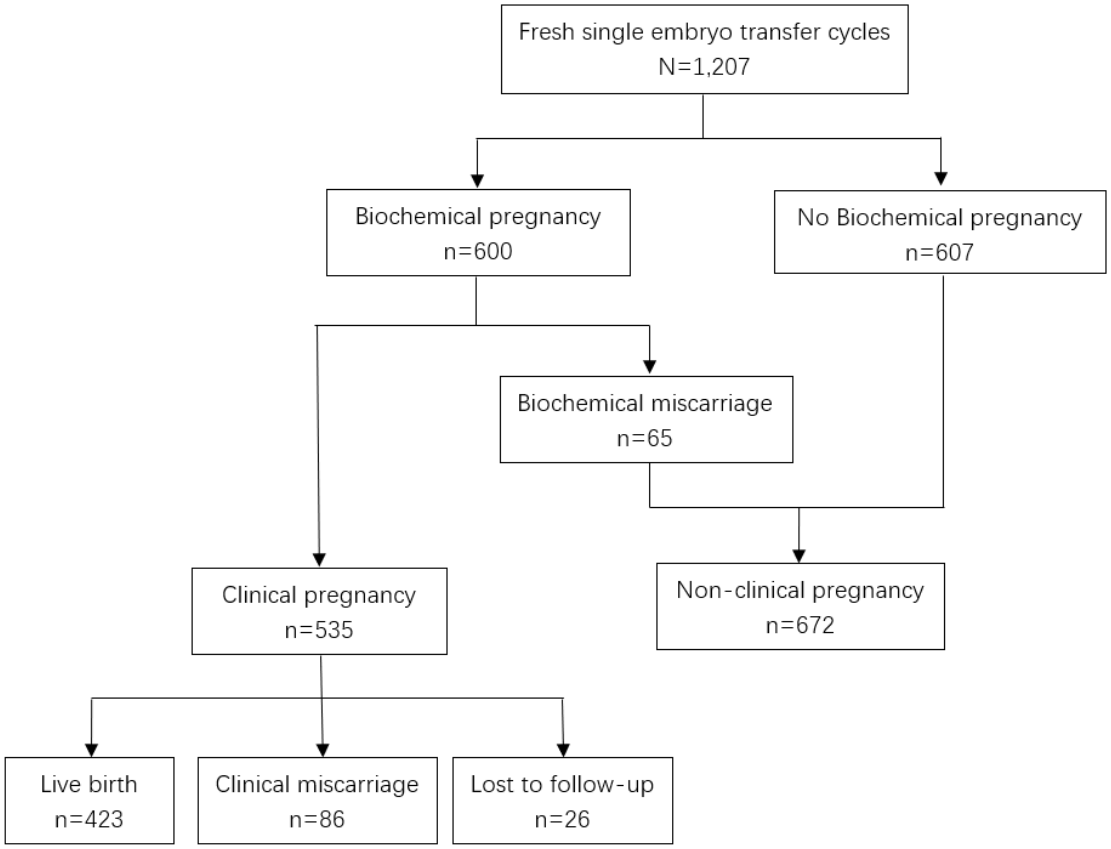


Figure 1. Flowchart of cycle details and pregnancy outcomes.

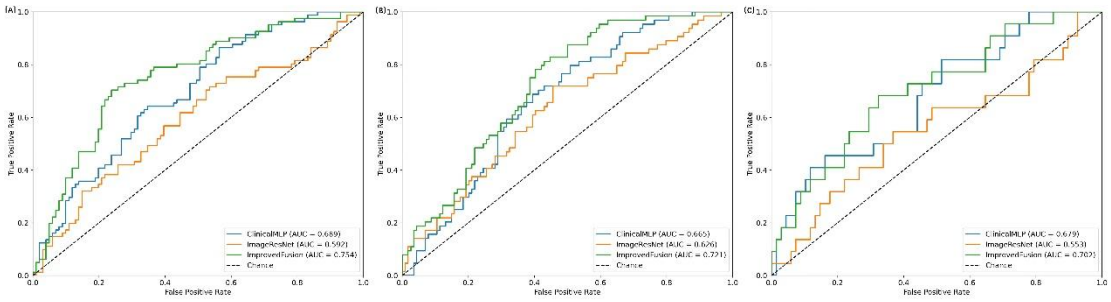


Figure 2. Performance of the three models in predicting pregnancy outcomes on the test set. (A) ROC curves for clinical pregnancy prediction. (B) ROC curves for live birth prediction. (C) ROC curves for miscarriage prediction. The blue, orange, and green curves represent the clinical-only, embryo image-only, and multimodal fusion models, respectively.

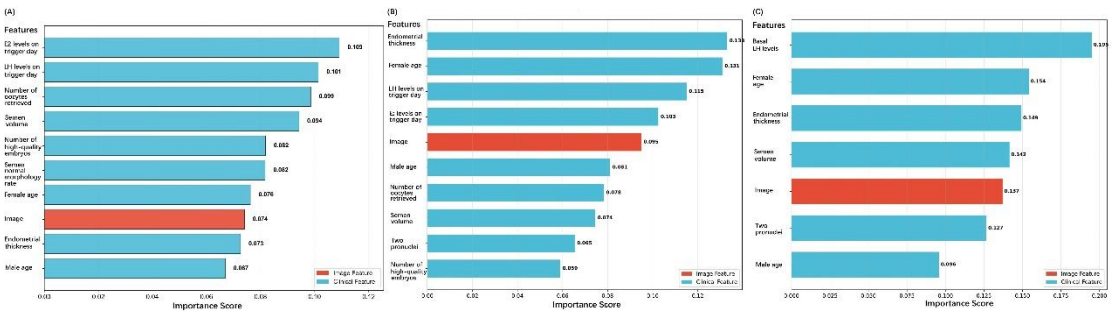


Figure 3. SHAP-based feature importance rankings for the ImprovedFusion model across the three prediction tasks. (A) Clinical pregnancy prediction. (B) Live birth prediction. (C) Miscarriage prediction. Bars represent the mean absolute SHAP values of the selected clinical features, indicating their relative contributions to each prediction outcome.

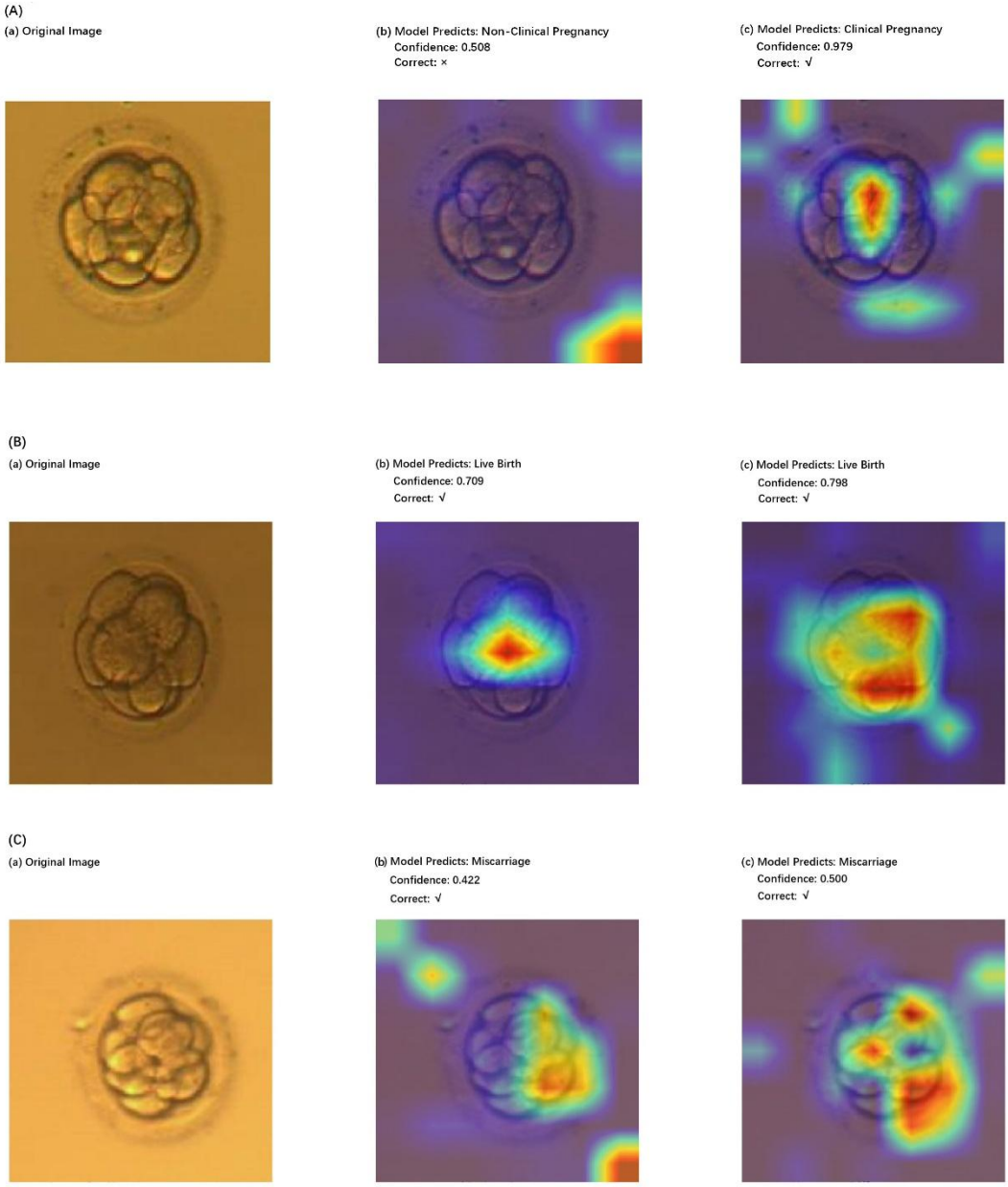


Figure 2. Visualization of Model Attention

(A) Clinical Pregnancy

(B) Live Birth

(C) Miscarriage

(D) Legend: Representative examples of D3 cleavage-stage embryos showing (a) original image, (b) heatmap from the ImageResNet model, and (c) heatmap from the ImprovedFusion model incorporating clinical features. Warmer colors indicate higher activation regions. The multimodal model demonstrates more focused attention on cell–cell interfaces, correlating with improved prediction confidence. Confidence scores represent the model's predicted probability for the given outcome, as visualized by the heatmap.

528 **Table 1. Baseline Characteristics of the Study Population Stratified by Clinical**
529 **Pregnancy Outcome**

Characteristic	Overall N = 1207[a]	CP Negative N = 672[a]	CP Positive N = 535[a]	p-value[b]
Age (years)	33.00 (30.00, 36.00)	34.00 (31.00, 37.00)	32.00 (30.00, 34.00)	<0.001
Male Age (years)	34.00 (31.00, 38.00)	35.00 (32.00, 39.00)	33.00 (31.00, 36.00)	<0.001
Infertility Type				0.002
Primary	586 (48.55%)	299 (44.49%)	287 (53.64%)	
Secondary	621 (51.45%)	373 (55.51%)	248 (46.36%)	
Duration of Infertility (years)	5.40 (4.00, 7.00)	5.40 (4.00, 7.00)	5.40 (4.00, 6.00)	0.231
Obstetric History				
Gravidity				0.005
0	529 (43.83%)	264 (39.29%)	265 (49.53%)	
1	316 (26.18%)	190 (28.27%)	126 (23.55%)	
2	217 (17.98%)	130 (19.35%)	87 (16.26%)	
≥3	145 (12.01%)	88 (13.10%)	57 (10.65%)	
Parity				0.096
No	848 (70.26%)	459 (68.30%)	389 (72.71%)	
Yes	359 (29.74%)	213 (31.70%)	146 (27.29%)	

Characteristic	Overall N = 1207[a]	CP Negative N = 672[a]	CP Positive N = 535[a]	p-value[b]
History of Miscarriage				0.001
No	781 (64.71%)	408 (60.71%)	373 (69.72%)	
Yes	426 (35.29%)	264 (39.29%)	162 (30.28%)	
History of Ectopic Pregnancy				0.830
No	1,025 (84.92%)	572 (85.12%)	453 (84.67%)	
Yes	182 (15.08%)	100 (14.88%)	82 (15.33%)	
Body Mass Index (kg/m²)	22.30 (20.30, 24.60)	22.50 (20.30, 24.80)	22.00 (20.20, 24.30)	0.066
Basal FSH (IU/L)	7.46 (6.21, 8.57)	7.61 (6.22, 8.85)	7.33 (6.19, 8.42)	0.074
Basal E_2 (pg/mL)	38.90 (28.90, 48.20)	38.80 (28.05, 48.72)	38.90 (29.50, 47.10)	0.939
Basal Prolactin (ng/mL)	18.58 (12.38, 25.13)	18.72 (11.93, 25.13)	18.17 (12.98, 25.13)	0.854
Basal LH (IU/L)	5.19 (3.86, 6.48)	5.16 (3.90, 6.35)	5.20 (3.78, 6.68)	0.883
Basal T (ng/dL)	0.25 (0.17, 0.30)	0.25 (0.16, 0.30)	0.25 (0.17, 0.31)	0.615
Semen Volume (mL)	3.30 (2.40, 4.30)	3.20 (2.40, 4.10)	3.40 (2.50, 4.40)	0.009
Semen pH	7.40 (7.40, 7.50)	7.40 (7.40, 7.50)	7.40 (7.40, 7.50)	0.101
Proportion of Progressively Motile Sperm (%)	45.20 (30.00, 59.00)	44.00 (29.00, 58.00)	46.00 (31.00, 60.00)	0.053
Sperm Density (million/mL)	34.55 (16.78, 57.75)	35.24 (16.44, 60.41)	32.98 (16.93, 55.59)	0.657

Characteristic	Overall N = 1207[a]	CP Negative N = 672[a]	CP Positive N = 535[a]	p-value[b]
Normal Morphology Rate (%)	7.07 (4.80, 9.00)	6.90 (4.60, 8.80)	7.07 (4.80, 9.70)	0.008
Ovarian Stimulation Protocol				<0.001
GnRH agonist	699 (57.91%)	355 (52.83%)	344 (64.30%)	
GnRH antagonist	455 (37.70%)	277 (41.22%)	178 (33.27%)	
Others	53 (4.39%)	40 (5.95%)	13 (2.43%)	
Serum Hormone Levels on the Trigger Day				
E_2 (pg/mL)	1,887.00 (1,203.00, 2,741.00)	1,783.50 (1,082.00, 2,689.50)	2,065.00 (1,359.00, 2,798.00)	<0.001
LH (IU/L)	1.57 (0.92, 2.66)	1.68 (1.00, 3.05)	1.40 (0.80, 2.41)	<0.001
Progesterone (ng/mL)	0.60 (0.38, 0.82)	0.59 (0.38, 0.83)	0.60 (0.39, 0.82)	0.455
Number of Oocytes Retrieved	9.00 (5.00, 13.00)	8.00 (4.00, 12.00)	10.00 (7.00, 13.00)	<0.001
Number of 2PN Fertilized Oocytes	5.00 (2.00, 8.00)	4.00 (2.00, 8.00)	6.00 (4.00, 9.00)	<0.001
Number of High-quality Embryos	4.00 (1.00, 6.00)	3.00 (1.00, 5.00)	4.00 (2.00, 6.00)	<0.001
Endometrial Thickness (mm)	11.20 (9.90, 13.00)	11.00 (9.50, 12.70)	12.00 (10.00, 13.20)	<0.001
Endometrial Pattern				0.190
Type A	1,100 (91.14%)	606 (90.18%)	494 (92.34%)	
Others	107 (8.86%)	66 (9.82%)	41 (7.66%)	

[a] Median (Q1, Q3); n (%)

[b] Wilcoxon rank sum test; Pearson's Chi-squared test

Abbreviations: CP: clinical pregnancy; BMI: body mass index; FSH: follicle-stimulating hormone; E_2: estradiol; LH: luteinizing hormone; T: testosterone; 2PN: two pronuclei.

Table 2. Performance Comparison of the Three Prediction Models on the Test Set.

Model	AUC (95% CI)	Balanced Accuracy	Sensitivity	Specificity	Precision	F1-Score
Clinical Pregnancy prediction						
ClinicalMLP	0.689 (0.613–0.765)	0.624	0.802	0.446	0.537	0.644
ImageResNet	0.592 (0.508–0.676)	0.509	0.790	0.228	0.451	0.674
ImprovedFusion	0.754 (0.683–0.826)	0.662	0.889	0.436	0.558	0.686
Live Birth prediction						
ClinicalMLP	0.665 (0.586-0.745)	0.640	0.656	0.623	0.494	0.564
ImageResNet	0.626 (0.540-0.711)	0.583	0.719	0.447	0.422	0.532
ImprovedFusion	0.721 (0.647-0.795)	0.682	0.750	0.614	0.522	0.615
Miscarriage prediction						
ClinicalMLP	0.679 (0.551–0.808)	0.614	0.727	0.500	0.320	0.444
ImageResNet	0.553 (0.404–0.703)	0.566	0.500	0.632	0.306	0.379
ImprovedFusion	0.702 (0.576–0.828)	0.651	0.727	0.574	0.356	0.478

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval.

SUPPLEMENTAL MATERIAL

Supplemental Table 1. Dataset Partition for Training, Validation, and Test Set

Outcome	Total (N)	Positive (N)	Negative (N)	Training (N)	Validation (N)	Test (N)
Clinical Pregnancy	1,207	535	672	845	181	181
Live Birth	1,181[a]	423	758	826	177	178
Miscarriage	600	151	449	420	90	90

[a] Twenty-six cases were lost to follow-up for live birth.

[*] The dataset was split using stratified sampling into training, validation, and test sets in a ratio of 7:1.5:1.5.

Supplemental Table 2. Baseline Characteristics of the Study Population Stratified by Live Birth Outcome

Characteristic	Overall N = 1,181[a]	LB Negative N = 758[a]	LB Positive N = 423[a]	p-value[b]
Age (years)	33.00 (30.00, 36.00)	33.00 (31.00, 37.00)	32.00 (29.00, 34.00)	<0.001
Male Age (years)	34.00 (31.00, 38.00)	35.00 (32.00, 39.00)	33.00 (31.00, 36.00)	<0.001
Infertility Type				0.005
Primary	572 (48.43%)	344 (45.38%)	228 (53.90%)	
Secondary	609 (51.57%)	414 (54.62%)	195 (46.10%)	
Duration of Infertility (years)	5.00 (4.00, 7.00)	5.40 (4.00, 7.00)	5.00 (4.00, 6.00)	0.093
Obstetric History				
Gravidity				0.008
0	515 (43.61%)	303 (39.97%)	212 (50.12%)	
1	312 (26.42%)	212 (27.97%)	100 (23.64%)	
2	212 (17.95%)	143 (18.87%)	69 (16.31%)	
≥3	142 (12.02%)	100 (13.19%)	42 (9.93%)	
Parity				0.056
No	828 (70.11%)	517 (68.21%)	311 (73.52%)	
Yes	353 (29.89%)	241 (31.79%)	112 (26.48%)	
History of Miscarriage				0.002

Characteristic	Overall N = 1,181[a]	LB Negative N = 758[a]	LB Positive N = 423[a]	p-value[b]
No	763 (64.61%)	465 (61.35%)	298 (70.45%)	
Yes	418 (35.39%)	293 (38.65%)	125 (29.55%)	
History of Ectopic Pregnancy				0.500
No	1,005 (85.10%)	649 (85.62%)	356 (84.16%)	
Yes	176 (14.90%)	109 (14.38%)	67 (15.84%)	
Body Mass Index (kg/m^2)	22.30 (20.30, 24.70)	22.50 (20.30, 24.80)	22.00 (20.20, 24.30)	0.102
Basal FSH (IU/L)	7.48 (6.21, 8.58)	7.57 (6.20, 8.79)	7.35 (6.22, 8.47)	0.292
Basal E_2(pg/mL)	38.90 (28.80, 48.31)	38.60 (28.19, 48.74)	39.11 (29.40, 47.10)	0.918
Basal Prolactin (ng/mL)	18.48 (12.38, 25.13)	18.61 (12.13, 25.13)	18.17 (13.00, 25.13)	0.759
Basal LH (IU/L)	5.19 (3.86, 6.48)	5.10 (3.85, 6.31)	5.32 (3.89, 6.87)	0.156
Basal T(ng/dL)	0.25 (0.17, 0.30)	0.25 (0.16, 0.30)	0.25 (0.17, 0.31)	0.218
Semen Volume (mL)	3.30 (2.40, 4.20)	3.20 (2.30, 4.10)	3.40 (2.50, 4.40)	0.003
Semen pH	7.40 (7.40, 7.50)	7.40 (7.40, 7.50)	7.40 (7.40, 7.50)	0.082
Proportion of Progressively Motile Sperm (%)	46.00 (30.00, 59.00)	45.00 (29.00, 58.00)	46.00 (31.00, 61.00)	0.096
Sperm Density (million/mL)	34.65 (16.93, 57.89)	35.39 (17.03, 59.36)	32.69 (16.74, 55.79)	0.479
Normal Morphology Rate (%)	7.07 (4.70, 9.00)	7.07 (4.70, 8.80)	7.07 (4.80, 9.70)	0.084

Characteristic	Overall N = 1,181[a]	LB Negative N = 758[a]	LB Positive N = 423[a]	p-value[b]
Ovarian Stimulation Protocol				<0.001
GnRH agonist	680 (57.58%)	409 (53.96%)	271 (64.07%)	
GnRH antagonist	449 (38.02%)	306 (40.37%)	143 (33.81%)	
Others	52 (4.40%)	43 (5.67%)	9 (2.13%)	
Serum Hormone Levels on the Trigger Day				
E_2(pg/mL)	1,883.00 (1,193.00, 2,729.00)	1,811.00 (1,109.00, 2,707.00)	2,022.00 (1,352.00, 2,767.00)	0.004
LH (IU/L)	1.57 (0.92, 2.66)	1.66 (0.99, 3.00)	1.40 (0.79, 2.34)	<0.001
Progesterone (ng/mL)	0.59 (0.38, 0.82)	0.59 (0.38, 0.82)	0.60 (0.39, 0.83)	0.485
Number of Oocytes Retrieved	9.00 (5.00, 13.00)	9.00 (4.00, 12.00)	10.00 (7.00, 13.00)	<0.001
Number of 2PN Fertilized Oocytes	5.00 (2.00, 8.00)	5.00 (2.00, 8.00)	6.00 (4.00, 9.00)	<0.001
Number of High-quality Embryos	3.00 (1.00, 6.00)	3.00 (1.00, 5.00)	4.00 (2.00, 6.00)	<0.001
Endometrial Thickness (mm)	11.20 (9.90, 13.00)	11.00 (9.50, 12.70)	12.00 (10.00, 13.40)	<0.001
Endometrial Pattern				0.578
Type A	1,076 (91.11%)	688 (90.77%)	388 (91.73%)	
Others	105 (8.89%)	70 (9.23%)	35 (8.27%)	

[a] Median (Q1, Q3); n (%)

[b] Wilcoxon rank sum test; Pearson's Chi-squared test

547 Abbreviations: LB: live birth; BMI: body mass index; FSH: follicle-stimulating hormone; E_2: estradiol; LH: luteinizing hormone; T:
548 testosterone; 2PN: two pronuclei.

549

550 **Supplemental Table 3. Baseline Characteristics of the Study Population Stratified by**
551 **Miscarriage Outcome**

Characteristic	Overall N = 600[a]	Miscarriage Negative N = 449[a]	Miscarriage Positive N = 151[a]	p-value[b]
Age (years)	32.00 (30.00, 35.00)	32.00 (30.00, 34.00)	33.00 (30.00, 36.00)	<0.001
Male Age (years)	33.00 (31.00, 36.50)	33.00 (31.00, 36.00)	35.00 (31.00, 38.00)	0.005
Infertility Type				0.736
Primary	321 (53.50%)	242 (53.90%)	79 (52.32%)	
Secondary	279 (46.50%)	207 (46.10%)	72 (47.68%)	
Duration of Infertility (years)	5.00 (4.00, 6.00)	5.40 (4.00, 6.00)	5.00 (4.00, 6.00)	0.779
Obstetric History				
Gravidity				0.820
0	296 (49.33%)	226 (50.33%)	70 (46.36%)	
1	144 (24.00%)	104 (23.16%)	40 (26.49%)	
2	100 (16.67%)	74 (16.48%)	26 (17.22%)	
≥3	60 (10.00%)	45 (10.02%)	15 (9.93%)	
Parity				0.400

Characteristic	Overall N = 600[a]	Miscarriage Negative N = 449[a]	Miscarriage Positive N = 151[a]	p-value[b]
No	437 (72.83%)	331 (73.72%)	106 (70.20%)	
Yes	163 (27.17%)	118 (26.28%)	45 (29.80%)	
History of Miscarriage				0.338
No	416 (69.33%)	316 (70.38%)	100 (66.23%)	
Yes	184 (30.67%)	133 (29.62%)	51 (33.77%)	
History of Ectopic Pregnancy				0.137
No	510 (85.00%)	376 (83.74%)	134 (88.74%)	
Yes	90 (15.00%)	73 (16.26%)	17 (11.26%)	
Body Mass Index (kg/m²)	22.00 (20.15, 24.35)	22.00 (20.20, 24.30)	22.00 (20.10, 24.90)	0.673
Basal FSH (IU/L)	7.36 (6.21, 8.47)	7.33 (6.23, 8.44)	7.51 (6.15, 8.55)	0.699
Basal E_2(pg/mL)	38.81 (29.35, 46.50)	39.11 (29.60, 46.60)	37.70 (28.80, 46.40)	0.541
Basal Prolactin (ng/mL)	18.96 (13.09, 25.13)	18.30 (13.02, 25.13)	19.30 (13.55, 26.31)	0.298
Basal LH (IU/L)	5.21 (3.79, 6.65)	5.32 (3.87, 6.74)	4.92 (3.59, 6.18)	0.041
Basal T(ng/dL)	0.25 (0.17, 0.31)	0.25 (0.17, 0.31)	0.22 (0.15, 0.30)	0.197
Semen Volume (mL)	3.30 (2.50, 4.35)	3.40 (2.50, 4.40)	3.10 (2.30, 4.00)	0.007
Semen pH	7.40 (7.40, 7.50)	7.40 (7.40, 7.50)	7.40 (7.40, 7.50)	0.547

Characteristic	Overall N = 600[a]	Miscarriage Negative N = 449[a]	Miscarriage Positive N = 151[a]	p-value[b]
Proportion of Progressively Motile Sperm (%)	47.00 (32.00, 60.00)	46.00 (31.00, 60.00)	47.00 (32.00, 59.00)	0.954
Sperm Density(million/mL)	32.89 (16.93, 55.79)	32.30 (16.64, 55.59)	35.53 (18.69, 57.63)	0.443
Normal Morphology Rate (%)	7.07 (4.80, 9.60)	7.07 (4.80, 9.70)	7.07 (5.50, 9.10)	0.811
Ovarian Stimulation Protocol				0.079
GnRH agonist	381 (63.50%)	290 (64.59%)	91 (60.26%)	
GnRH antagonist	200 (33.33%)	149 (33.18%)	51 (33.77%)	
Others	19 (3.17%)	10 (2.23%)	9 (5.96%)	
Serum Hormone Levels on the Trigger Day				
E_2(pg/mL)	2,038.50 (1,333.50, 2,775.00)	2,022.00 (1,369.00, 2,770.00)	2,048.00 (1,107.00, 2,812.00)	0.546
LH (IU/L)	1.44 (0.82, 2.45)	1.40 (0.79, 2.34)	1.53 (0.92, 2.56)	0.131
Progesterone (ng/mL)	0.60 (0.39, 0.83)	0.60 (0.39, 0.83)	0.60 (0.39, 0.82)	0.723
Number of Oocytes Retrieved	10.00 (7.00, 13.00)	10.00 (7.00, 13.00)	9.00 (6.00, 13.00)	0.069
Number of 2PN Fertilized Oocytes	6.00 (4.00, 9.00)	6.00 (4.00, 9.00)	6.00 (3.00, 8.00)	0.040
Number of High-quality Embryos	4.00 (2.00, 6.00)	4.00 (3.00, 6.00)	4.00 (2.00, 6.00)	0.195
Endometrial Thickness (mm)	11.70 (10.00, 13.00)	12.00 (10.00, 13.40)	11.00 (10.00, 12.60)	0.003

Characteristic	Overall N = 600[a]	Miscarriage Negative N = 449[a]	Miscarriage Positive N = 151[a]	p-value[b]
Endometrial Pattern				0.235
Type A	555 (92.50%)	412 (91.76%)	143 (94.70%)	
Others	45 (7.50%)	37 (8.24%)	8 (5.30%)	

[a] Median (Q1, Q3); n (%)

[b] Wilcoxon rank sum test; Pearson's Chi-squared test

Abbreviations: BMI: body mass index; FSH: follicle-stimulating hormone; E_2: estradiol; LH: luteinizing hormone; T: testosterone; 2PN: two pronuclei.

Supplemental Table 4. Clinical Features Included in the Prediction Models by Outcome

Demographics	Feature	CP	LB	Miscarriage
Demographics				
	Age (years)	+	+	+
	Male Age (years)	+	+	+
	Gravidity	+	+	
	Parity	+	+	
Cycle Characteristics				
	Infertility Type	+	+	
	Ovarian Stimulation Protocol	+	+	
	Number of Oocytes Retrieved	+	+	
	Number of 2PN Fertilized Oocytes	+	+	+
	Number of High-quality Embryos	+	+	
	Endometrial Thickness (mm)	+	+	+
	Endometrial Pattern			
Semen Parameters				
	Semen Volume (mL)	+	+	+
	Normal Morphology Rate (%)	+		
Hormonal Profile				
	Basal LH (IU/L)			+
	E_2 Levels on the Trigger Day (pg/mL)	+	+	
	LH Levels on the Trigger Day (IU/L)	+	+	

Abbreviations: CP: clinical pregnancy; LB: live birth; 2PN: two pronuclei; LH: luteinizing hormone; E_2: estradiol.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.docx
- SupplementaryTable2.docx
- SupplementaryTable3.docx
- SupplementaryTable4.docx