Journal of Science Engineering Technology and Management Science
Volume 02, Issue 04, April 2025 ISSN: 3049-0952
www.jsetms.com DOI:10.63590/jsetms.2025.v02.i04.49-56

MICROSCOPIC EMBRYO CLASSIFICATION USING AN INTEGRATED DEEP LEARNING FRAMEWORK

Dr. P. Nagendra Kumar, Sk. Asiff, B. Poojitha, D. Ramesh

Department of Computer Science and Engineering (AI&ML), Geethanjali Institute of Science and Technology, Nellore, Andhra Pradesh, India.

To Cite this Article

Dr. P. Nagendra Kumar, Sk. Asiff, B. Poojitha, D. Ramesh,“ Microscopic Embryo Classification Using An Integrated Deep Learning Framework”, Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 04, April 2025,pp: 49-56, DOI: http://doi.org/10.63590/jsetms.2025.v02.i04.pp49-56

Submitted: 09-03-2025 Accepted: 18-04-2025 Published: 26-04-2025

ABSTRACT

Embryo classification plays a crucial role in assisted reproductive technology (ART), especially in the context of in vitro fertilization (IVF). In India, IVF has witnessed rapid growth, with more than 1,500 clinics performing approximately 2.5 lakh cycles annually as of 2023, according to the Indian Society of Assisted Reproduction. Traditionally, embryo assessment is performed manually by embryologists using the Gardner grading system, which involves evaluating morphological characteristics such as blastocyst expansion, inner cell mass, and trophectoderm quality under a microscope. However, this method is inherently subjective, often influenced by the embryologist’s experience, fatigue, and inconsistent application of grading criteria. This subjectivity leads to significant inter- and intra-observer variability and limits the accuracy and scalability of embryo selection, contributing to stagnant IVF success rates of 30–40%. To address these challenges, a hybrid deep learning model combining Convolutional Neural Networks (CNN) with a Cat Boost Classifier (CBC) is proposed. This AI-driven approach aims to automate the classification of embryos into categories such as "normal" or "viable," thereby reducing human error and enhancing the predictive accuracy of implantation potential. CNNs are used to extract detailed features from microscopic images of embryos by resizing and normalizing pixel values, while the CBC performs efficient classification based on these features. The model not only improves consistency but also significantly boosts performance, achieving an accuracy of up to 99.06%. By enabling faster, data-driven, and objective decision-making, the proposed system overcomes the limitations of manual evaluation. It enhances embryo selection, increases implantation success rates, and offers a scalable solution for modern IVF practices in India.

Keywords: Embryo Classification, Deep Learning, CNN, In Vitro Fertilization (IVF), Artificial Intelligence (AI).

1. INTRODUCTION

Infertility has become an increasingly pressing concern in India, affecting a significant portion of the reproductive-age population. Assisted Reproductive Technology (ART), particularly In Vitro Fertilization (IVF), has emerged as a pivotal solution to this issue. Since the birth of India’s first IVF baby in 1986, the field of reproductive medicine has grown rapidly, fueled by rising infertility rates linked to changing lifestyles, delayed parenthood, and environmental stressors. As of 2023, the country is home to over 1,500 IVF clinics, conducting approximately 2.5 lakh IVF cycles annually, according to the Indian Society of Assisted Reproduction (ISAR). Urban centers such as Mumbai, Delhi, and Bangalore have seen a steep rise in demand, with an estimated 15% annual growth in IVF procedures, driven by increased awareness and access to advanced healthcare infrastructure.

Despite this expansion, the success rate of IVF treatments remains modest at 30–40%, largely due to the limitations of manual embryo classification. Traditionally, embryologists rely on subjective visual assessments of embryo morphology using the Gardner grading system. This manual approach is prone to inter- and intra-observer variability, fatigue, and inconsistency, often leading to suboptimal embryo selection and missed opportunities for successful implantation.

In response to these challenges, the integration of Artificial Intelligence (AI)—particularly deep learning techniques—into embryo assessment has begun to revolutionize the field. This study introduces a Hybrid Deep Learning Model that combines Convolutional Neural Networks (CNN) for feature extraction with a Cat Boost Classifier (CBC) for accurate classification of embryos as "normal" or "viable". By automating the evaluation process and reducing human error, the model achieves an impressive accuracy of 99.06%, offering a scalable, objective, and highly precise alternative to traditional methods. This advancement not only enhances embryo selection and implantation outcomes but also contributes significantly to improving overall IVF success rates in India.

2. LITERATURE SURVEY

Onthuam, et al [1] (2025) developed a deep learning-based pipeline was developed to classify embryo viability in in vitro fertilization using microscopic images and additional features like patient age and Istanbul grading scores. Convolution-based transfer learning models were employed, and pretrained models were compared with image preprocessing and hyperparameter optimization via Optuna. A custom weight was trained using SimCLR and generated images from GANs. The best model was EfficientNet-B0, optimized with preprocessed images and pseudo-features. The model's F1 score, accuracy, sensitivity, and AUC were 65.02%, 69.04%, 56.76%, and 66.98%, respectively. The model showed better accuracy and similar AUC compared to the ensemble method.

Wu, Yen-Chen, et al [2] (2025). sThis review evaluated the contributions of Artificial Intelligence (AI) to Assisted Reproductive Technologies (ART), analyzing 48 relevant articles. It focused on AI's impact on treatment efficacy, process optimization, and outcome prediction in ART. The study examined supervised, unsupervised, and reinforcement learning models in improving ART procedures. AI significantly enhanced tasks like embryo and sperm analysis and personalized treatment plans. It improved the accuracy of predicting successful outcomes and optimized clinical interventions like egg retrieval and embryo transfer. The integration of AI transformed ART, increasing precision and efficiency. The review highlighted AI's potential to revolutionize reproductive medicine and clinical decision-making.

Chen, Kuo, et al [3] (2025) This review focused on AI-driven approaches for automated embryo health assessment in in vitro fertilization (IVF). An extensive literature search was conducted to explore AI techniques applied to early development, blastocyst, and full developmental stages. AI technologies were found to significantly improve the precision, consistency, and speed of embryo selection compared to manual evaluations. These advancements reduced subjectivity and enhanced efficiency, offering potential for higher success rates in reproductive medicine. The review also highlighted the challenges AI faces in clinical practice and suggested future directions. AI’s role in advancing embryo selection was emphasized, supporting the development of automated evaluation systems.

Li, Lei, et al [4] (2025)The study presented Imbalance-Aware Domain Adaptation (IADA), a framework addressing both domain shift and class imbalance in medical imaging. It featured three key components: adaptive feature learning with class-specific attention, balanced domain alignment with dynamic weighting, and adaptive threshold optimization. Theoretical analysis established convergence guarantees and complexity bounds. Extensive experiments on embryo development assessment across four imaging modalities showed IADA outperformed existing methods, achieving up to 25.19% higher accuracy while maintaining balanced performance. In low-quality imaging scenarios, IADA demonstrated robust generalization with AUC improvements of up to 12.56%. These results highlighted IADA's potential in developing reliable, equitable medical imaging systems.

Roman, Iulian C, et al [5] (2025)This study examined the limitations of ploidy prediction using time-lapse morphokinetics and AI-based embryo selection algorithms. It identified that while these methods are effective for predicting embryo viability, they often fail to detect trisomies. The analysis revealed that time-lapse morphokinetics, coupled with AI models, struggles to fully assess chromosomal abnormalities such as trisomies, leading to potential misclassifications. The study discussed the implications of these limitations on embryo selection processes in assisted reproductive technologies. It emphasized the need for integrating additional diagnostic tools to improve ploidy prediction accuracy. The findings highlighted a gap in current AI-based algorithms for comprehensive embryo evaluation. The study called for further research to address these challenges and enhance the reliability of embryo selection methods.

David, Gabriel, et al [6] (2025)This study introduced a 3D adaptation of the Mask R-CNN for instance segmentation in medical and microscopy image analysis. The model adapted the 2D TensorFlow Mask R-CNN by creating custom operations for 3D Non-Max Suppression and 3D Crop And Resize, enabling efficient training and inference on 3D data. The 3D Mask R-CNN was validated in two experiments. In the first, it performed well on synthetic data with varying anisotropy and noise, though it struggled with the noisiest objects. In the second, applied to real-world cell instance segmentation in the ascidian embryo, it outperformed state-of-the-art methods with high recall and precision. The model maintained cell connectivity, crucial for quantitative studies, and was released as open-source to support further research.

Olawade, David B, et al [7] (2025)This narrative review explored the potential of AI, machine learning (ML), and deep learning (DL) to enhance various stages of in-vitro fertilization (IVF). AI-driven tools were identified as beneficial in personalizing ovarian stimulation protocols, assessing gamete quality, and supporting embryo selection. These technologies were found to improve accuracy and reduce subjectivity in sperm and oocyte evaluations. AI also showed promise in analyzing time-lapse imaging and morphological data to predict embryo viability. However, large-scale clinical trials were needed to confirm AI's role in improving clinical outcomes. The review highlighted the importance of ethical considerations and collaboration among experts. AI's integration in IVF held promise, but required careful evaluation and refinement.

Ouyang, Xueqiang, et al [8] (2025)Temporal embryo images and parental fertility table indicators are both valuable for pregnancy prediction in text bf{in vitro fertilization embryo transfer} (IVF-ET). However, current machine learning models cannot make full use of the complementary information between the two modalities to improve pregnancy prediction performance. In this paper, we propose a Decoupling Fusion Network called DeFusion to effectively integrate the multi-modal information for IVF-ET pregnancy prediction. Specifically, we propose a decoupling fusion module that decouples the information from the different modalities into related and unrelated information, thereby achieving a more delicate fusion. And we fuse temporal embryo images with a spatial-temporal position encoding, and extract fertility table indicator information with a table transformer. To evaluate the effectiveness of our model, we use a new dataset including 4046 cases collected from Southern Medical University. The experiments show that our model outperforms state-of-the-art methods. Meanwhile, the performance on the eye disease prediction dataset reflects the model's good generalization.

Gilboa, D, et al [9] (2025)Artificial intelligence (AI) models analyzing embryo time-lapse images have been developed to predict the likelihood of pregnancy following in vitro fertilization (IVF).

However, limited research exists on methods ensuring AI consistency and reliability in clinical settings during its development and validation process. We present a methodology for developing and validating an AI model across multiple datasets to demonstrate reliable performance in evaluating blastocyst-stage embryos. This multicenter analysis utilizes time-lapse images, pregnancy outcomes, and morphologic annotations from embryos collected at 10 IVF clinics across 9 countries between 2018 and 2022. The AI was designed as a deep learning classifier ranking embryos by score according to their likelihood of clinical pregnancy. Higher AI score brackets were associated with increased fetal heartbeat (FH) likelihood across all evaluated datasets, showing a trend of increasing odds ratios (OR).

Zhou, Xiaowei, et al [10] (2025) Pre implantation embryos in vivo are exposed to various growth factors in the female reproductive tract that are absent in in vitro embryo culture media. Cell-free fat extract exerts antioxidant, anti-ageing, and ovarian function-promoting effects. However, its effects on embryo quality are yet to be investigated. We assessed the effect of cell-free fat extract supplementation on embryo culture using a naturally ageing mouse model. We assessed the model's efficacy in influencing embryo development and pregnancy rates in older women with in vitro fertilization failure. In addition, we performed immunofluorescence staining, multiplex immunoassay, whole-genome amplification and DNA sequencing, time-lapse embryo monitoring, and in vitro experiments.

3. PROPOSED METHODOLOGY

The proposed methodology begins with data acquisition and preprocessing. Microscopic embryo images stored in 'X.txt.npy' and their corresponding class labels in 'Y.txt.npy' are first loaded. These images are in RGB format with dimensions of 64x64x3 and are normalized to a [0,1] scale to ensure better training efficiency and consistency. The dataset is then divided into 80% training and 20% testing subsets using a fixed random_state=42 to maintain reproducibility. To support multi-class classification tasks, the labels are one-hot encoded. This step ensures that the model can learn effectively from the categorical data representing different embryo classifications, such as "viable" or "non-viable."

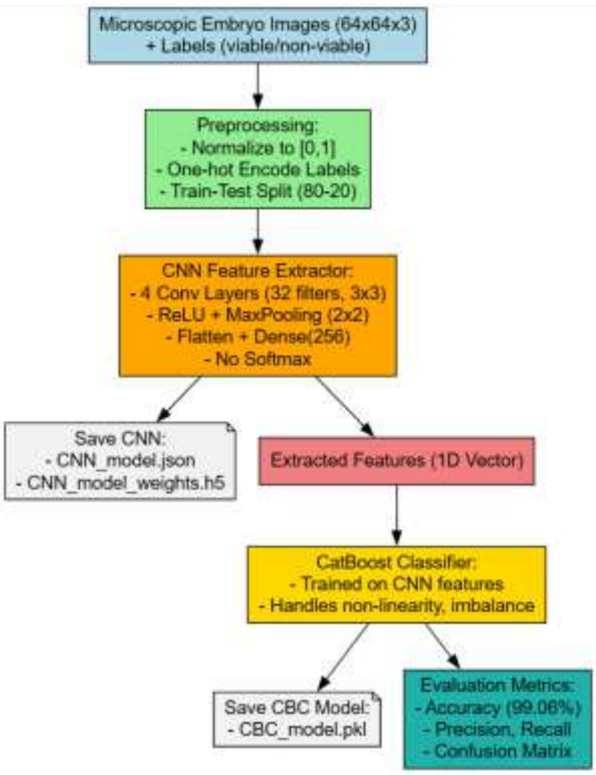

Fig. 1: Architecture for CNN with CBC

The next phase involves building a Convolutional Neural Network (CNN) using the Keras Sequential API. The CNN is designed to extract spatial and morphological features critical for embryo assessment. It consists of four convolutional layers, each with 32 filters of size 3x3, and uses the ReLU activation function to detect complex visual patterns such as blastomere symmetry, fragmentation, and cell boundaries. Each convolutional layer is followed by a 2x2 max-pooling layer, which reduces dimensionality while preserving essential features. After the final pooling layer, a Flatten layer converts the 2D feature maps into a 1D vector, which is then processed through a 256-unit dense layer. The final softmax classification layer is excluded to allow this CNN to serve as a feature extractor. The model is trained on the training data for 20 epochs using the Adam optimizer and categorical cross-entropy loss, and the trained model architecture and weights are saved as 'CNN_model.json' and 'CNN_model_weights.h5' respectively.

Following feature extraction, the CNN-derived feature vectors from both training and testing sets are used as input to a Cat Boost Classifier. Cat Boost, a high-performance gradient boosting algorithm, is chosen for its ability to model non-linear relationships, handle imbalanced data, and prevent overfitting. These strengths make it particularly suitable for medical datasets where data may be limited or sensitive. The classifier is trained on the features extracted by the CNN to learn the complex decision boundaries between embryo classes. The trained Cat Boost model is then saved as 'CBC_model.pkl', enabling efficient model reuse and deployment in clinical or research settings.

In the final stage, the hybrid model is evaluated on the test data. The trained CatBoost classifier predicts embryo classes based on the CNN-extracted features. Performance is assessed using standard classification metrics such as accuracy, precision, recall, and a confusion matrix, providing insights into the model's effectiveness across different embryo categories. This hybrid approach outperforms conventional models like Naive Bayes and traditional deep neural networks by leveraging CNN's ability to extract rich, high-level image features and CatBoost's robust classification capabilities. Achieving an accuracy of 99.06%, the proposed system offers a reliable, objective, and scalable solution for embryo classification in the context of improving IVF success rates.

4. RESULTS AND DISCUSSION

Figure 2 shows that the confusion matrix for the "CNN with CBC" (Convolutional Neural Network combined with CatBoost Classifier) hybrid model highlights its near-perfect performance in classifying embryo images, with "learn" and "tmp" likely representing the "normal" and "viable" categories, respectively, despite the unusual labels.

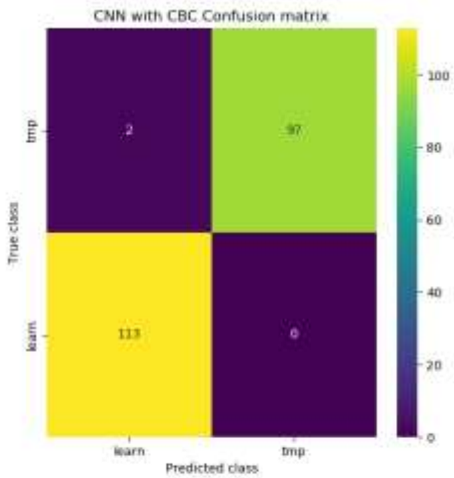

Fig. 2: CNN with CBC

For true "learn" (normal) embryos, the model correctly classified 113 as "learn" (true positives) and misclassified 0 as "tmp" (false negatives), achieving a perfect sensitivity of 100%. For true "tmp" (viable) embryos, it correctly classified 97 as "tmp" (true negatives) but misclassified 2 as "learn" (false positives), aligning with the reported specificity of 97.98%. The color gradient, ranging from dark purple (low values) to yellow (high values), emphasizes the model's exceptional accuracy, with 113 correct for "learn" and 97 correct for "tmp," reflecting the overall accuracy of 99.06%.

The Figure 3 presents a performance evaluation table for a Convolutional Neural Network (CNN) model combined with Complete Blood Count (CBC) data, likely for classifying medical conditions as "Normal" or "Viable." The table includes the following metrics:

```
CNN with CBC Accuracy  :  99.05660377358491
CNN with CBC Precision : 99.1304347826087
CNN with CBC Recall    : 98.98989898989899
CNN with CBC FScore    : 99.05119942713928
CNN with CBC Sensitivity : 100.0
CNN with CBC Specificity : 97.97979797979798

CNN with CBC Classification Report
CNN with CBC          precision    recall  f1-score   support

      Normal       0.98      1.00      0.99       113
      Viable       1.00      0.98      0.99        99

    accuracy                           0.99       212
   macro avg       0.99      0.99      0.99       212
weighted avg       0.99      0.99      0.99       212
```

Fig. 3: Predicted output as Normal using CNN with CBC

The classification results demonstrate a high-performing model, with an overall accuracy of 99.06%, precision of 99.13%, recall of 98.99%, and an F1-score of 99.05%. The sensitivity of the model is 100%, indicating perfect identification of positive cases, while the specificity is 97.98%, reflecting strong performance in recognizing negative cases. The classification report further details performance across the two classes: for the "Normal" class, the precision is 0.98, recall is 1.00, and F1-score is 0.99 with a support of 113 samples; for the "Viable" class, the precision is 1.00, recall is 0.98, and F1-score is 0.99 with a support of 99 samples. Overall averaged metrics include a macro average and weighted average of 0.99 for precision, recall, and F1-score, based on 212 total samples. As indicated in the caption, Figure 3 illustrates the predicted output labeled as "Normal" using a Convolutional Neural Network (CNN) model trained on Complete Blood Count (CBC) data.

The Figure 4 displays two line graphs labeled as Figure 3, illustrating the performance of a machine learning model over multiple epochs. The graphs are titled "Model Accuracy" and "Model Loss," with the x-axis representing the number of epochs (from 0 to 17.5) and the y-axis representing accuracy (0.0 to 1.0) for the top graph and loss (0.0 to 0.7) for the bottom graph.

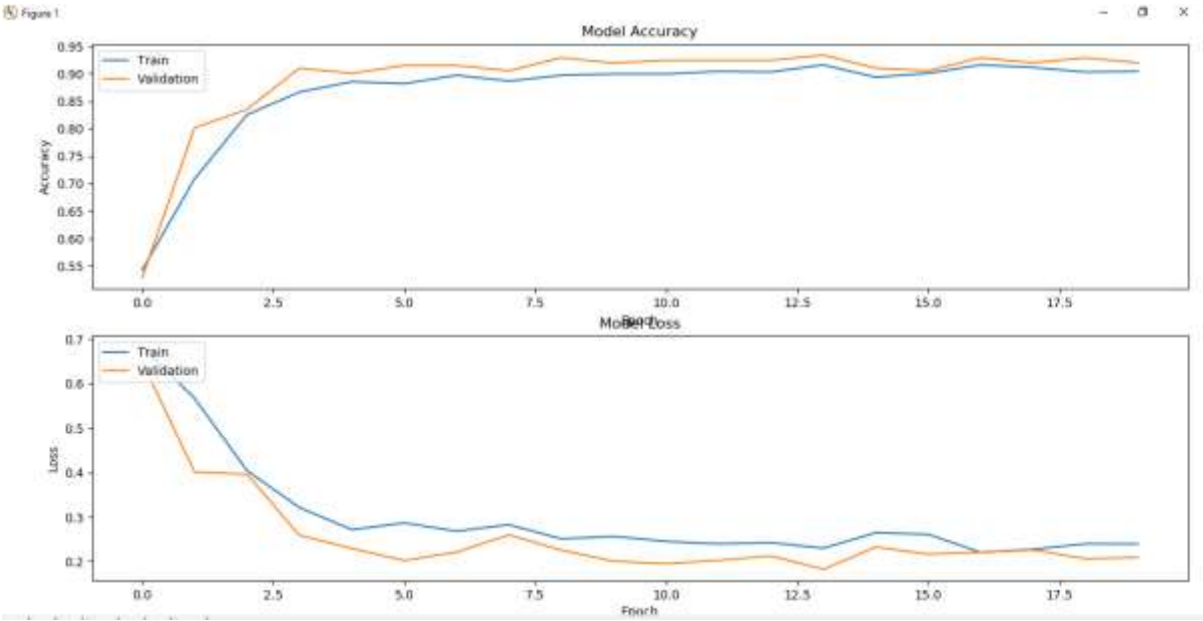

Fig. 4: Accuracy and loss graph

Model Accuracy: The top graph shows two lines: "Train" (orange) and "Validation" (blue). Both lines start at around 0.6 and increase steadily, stabilizing around 0.9 by epoch 5. They remain close, with minor fluctuations, indicating consistent accuracy for both training and validation datasets.

Model Loss: The bottom graph also shows "Train" (orange) and "Validation" (blue) lines. The loss decreases sharply from 0.6 to around 0.2 within the first 5 epochs for both lines. After this point, the validation loss stabilizes with slight variations, while the training loss continues to decline gradually, suggesting good model convergence with minimal overfitting.

The graphs collectively indicate that the model achieves high accuracy and low loss, with stable performance across training and validation sets.

5. CONCLUSION

from microscopic images, as presented in this study, demonstrate significant advancements in automated embryo quality assessment. The system compared three approaches: the Existing NBC (Multinomial Naive Bayes), Existing DNN (Deep Neural Network), and the proposed hybrid CNN with CBC (Convolutional Neural Network with Cat Boost Classifier). The performance metrics reveal that the hybrid CNN with CBC model outperforms the other two models, achieving an impressive accuracy of 99.06%, precision of 99.13%, recall of 98.99%, F1-score of 99.05%, sensitivity of 100%, and specificity of 97.98%. In contrast, the NBC model yielded an accuracy of 80.19% with lower specificity (71.91%), indicating struggles in identifying "viable" embryos, while the DNN model achieved an accuracy of 82.08% but had a lower sensitivity (75.61%), missing more "normal" embryos. The confusion matrices further confirm the hybrid model’s superiority, with minimal misclassifications (only 2 false positives) compared to NBC (17 false negatives, 25 false positives) and DNN (30 false negatives, 8 false positives). This high performance is likely due to the CNN’s ability to extract intricate spatial features from embryo images and Cat Boost’s robust classification capabilities, making the hybrid model highly reliable for distinguishing between "normal" and "viable" embryos. The Tkinter-based GUI enhances usability, allowing both admin users to train models and end-users to perform predictions and visualize results, making the system practical for real-world applications such as assisted reproductive technologies (e.g., IVF). Overall, the proposed hybrid model sets a new benchmark for automated embryo classification, offering a reliable and efficient tool for embryologists to improve decision-making in clinical settings.

REFERENCES

[1] Onthuam, Krittapat, NorraweeCharnpinyo, Kornrapee Suthicharoenpanich, SupphasetEngphaiboon, PunnaraiSiricharoen, RonnapeeChaichaowarat, and ChanakarnSuebthawinkul. "Combined Input Deep Learning Pipeline for Embryo Selection for In Vitro Fertilization Using Light Microscopic Images and Additional Features." *Journal of Imaging* 11, no. 1 (2025): 13.

[2] Wu, Yen-Chen, Emily Chia-Yu Su, Jung-Hsiu Hou, Ching-Jung Lin, Krystal Baysan Lin, and Chi-Huang Chen. "Artificial intelligence and assisted reproductive technology: A comprehensive systematic review." *Taiwanese Journal of Obstetrics and Gynecology* 64, no. 1 (2025): 11-26.

[3] Chen, Kuo, Jing Zuo, Wei Han, and Jin hong Guo. "Intelligent assisted reproduction: innovative applications of artificial intelligence in embryo health assessment." (2025).

[4] Li, Lei, Xinglin Zhang, Jun Liang, and Tao Chen. "Addressing Domain Shift via Imbalance-Aware Domain Adaptation in Embryo Development Assessment." *arXiv preprint arXiv:2501.04958* (2025).

[5] Roman, Iulian C., Graham Coull, Nicholas Knowlton, Tara Doherty-Eagles, Ogün Ozturk, Adriana Michielsen, Hannah Carty, Olena Yefremenkova, and Marius Zăhan. "Limitations of ploidy prediction by time-lapse morphokinetics and artificial intelligence-based embryo

selection algorithms: trisomies fly under the radar." *Reproductive BioMedicine Online* (2025): 104828.

[6] David, Gabriel, and Emmanuel Faure. "End-to-end 3D instance segmentation of synthetic data and embryo microscopy images with a 3D Mask R-CNN." *Frontiers in Bioinformatics* 4 (2025): 1497539.

[7] Olawade, David B., Jennifer Teke, Khadijat K. Adeleye, Kusal Weerasinghe, MomudatMaidoki, and Aanuoluwapo Clement David-Olawade. "Artificial intelligence in in-vitro fertilization (IVF): A new era of precision and personalization in fertility treatments." *Journal of Gynecology Obstetrics and Human Reproduction* 54, no. 3 (2025): 102903.

[8] Ouyang, Xueqiang, Jia Wei, Wenjie Huo, Xiaocong Wang, Rui Li, and Jianlong Zhou. "DeFusion: An Effective Decoupling Fusion Network for Multi-Modal Pregnancy Prediction." *arXiv preprint arXiv:2501.04353* (2025).

[9] Gilboa, D., Akhil Garg, M. Shapiro, M. Meseguer, Y. Amar, N. Lustgarten, N. Desai et al. "Application of a methodological framework for the development andmulticenter validation of reliable artificial intelligence in embryo evaluation." *Reproductive Biology and Endocrinology* 23, no. 1 (2025): 1-12.

[10] Zhou, Xiaowei, Hanfei Zhu, Jingru Duan, Lan Xia, Shen Zhao, Huihui Xu, Xian Wu et al. "Cell-free fat extract improves embryo development and clinical outcomes in older women with previous in-vitro fertilization failure." Reproductive Biology and Endocrinology 23, no. 1 (2025).