Proposal: Hybrid Real-Synthetic Blastocyst Classification

Ioannis Kasionis
Nikolas Kavaklis

June 16, 2025

Abstract

Manual grading of day-4 and day-5 human blastocysts by embryologists is subjective, prone to lab bias and require highly trained doctors. Machine classification is limited by small and unbalanced datasets due to the nature of the subject. We propose an end-to-end pipeline combining real-image classifiers, synthetic image generation, and explainability to improve grading consistency and downstream viability prediction. Key components include benchmarking against existing deep-learning baselines, exploring multiple generative approaches, and rigorous quantitative evaluation.

1 Introduction

Embryo selection in IVF relies heavily on morphological assessment of blastocyst images, scored by Gardner's criteria (Expansion, Inner Cell Mass, Trophectoderm). This process is subjective and suffers from inter-annotator variability and lab bias. Recent works on visual transformers like SWIN[3] and CNN like models have shown promise using deep learning on static images. Standardization of training and testing Datasets[1] have given a baseline and quality data for machine learning but are limited in size and suffer from uneven class distribution. Synthetic data generation offers a route to enrich underrepresented classes and boost classifier robustness[4].

2 Objectives

- Develop a multiple classifier's to predict Gardner's EXP, ICM, TE grades from single static images.
- Generate high-fidelity synthetic blastocyst images conditioned on all Gardner grades.
- Augment the real dataset with synthetic minority-class samples and retrain classifiers to compare with original results.
- Compare results against established baselines from Kromp et al. (2023)[1].

3 Dataset and Baseline Models

We will use the Annotated Human Blastocyst Dataset (Kromp et al. 2023)[1], comprising 2344 day-4 and day-5 images annotated with Gardner EXP (0–5), ICM (A–C), TE (A–C). The dataset is split to a "Silver" training set of 2044 images and 300 "Golden" test set. The original work benchmarks human-expert agreement and deep-learning baselines including:

- **XCeption**
- **Deit transformer**
- **Swin transformer**

4 Methodology

Our pipeline has two complementary threads:

4.1 Real-Image Classification

- Train multiple vision transformers(SWIN)[3] on the real images to predict EXP, ICM, TE grades. Train a model for each of the criteria for both model architectures.
- Evaluate using accuracy, macro-F1.
- Compare to Kromp et al. 2023)[1] deep-learning baselines.

4.2 Synthetic Image Generation

We will explore generative paradigms to synthesize realistic, label-conditioned blastocysts:

- **Conditional GANs** (e.g. StyleGAN2-ADA)

These methods will be compared on FID/KID.

4.3 Data Fusion and Re-training

- Identify minority classes and generate synthetic images to achieve balance.
- Retrain classifiers on the merged real + synthetic dataset.
- Measure improvements in macro-F1, minority-class recall, and overall κ.

5 Timeline and Resources

- **Step 1:** Data preprocessing, baseline reproduction, real-image classifier prototypes.
- **Step 2:** Synthetic model prototyping (GANs), initial FID/KID evaluation.
- **Step 3:** Data fusion, re-training classifiers, explainability integration.

References

[1] F. Kromp *et al.*, "An annotated human blastocyst dataset to benchmark deep learning architectures for in vitro fertilization," *Scientific Data*, vol. 10, 271, 2023.

[2] F. Chollet,"Xception: Deep Learning with Depthwise Separable Convolutions" *Google Inc* 2017

[3] Ze Liu *et al.*,"Swin Transformer: Hierarchical Vision Transformer using Shifted Windows"

[4] O. Presacan *et al.*"Merging synthetic and real embryodata for advanced AI predictions" *nature portfolio*