

Classification of Embryo Images for Viability Detection

Ester Molinari

September 16, 2025

Abstract

This report details the development of a computer vision model for the classification of pig embryo images to determine their viability. The primary challenge addressed was the severely imbalanced dataset, where viable embryos were significantly underrepresented. To mitigate this, a data augmentation strategy was implemented using transformations like random flips and rotations. We selected and trained two main architectures based on ResNet18, one incorporating a Long Short-Term Memory (LSTM) layer for classifying image sequences and a simpler version for single images. The models demonstrated strong performance, with the ResNet18 with LSTM model achieving a test accuracy of 96.62% on the augmented dataset. Our findings confirm that computer vision is an effective tool for embryo viability detection and that data augmentation is critical for building robust models on imbalanced biological datasets.

1 Introduction

The objective of this project was to develop and train a computer vision model capable of classifying images and sequences of images of pig embryos. The goal of this classification is to accurately assess the **viability of the embryos**. This is a crucial task for modern swine reproductive management programs, as effective pregnancy diagnostic procedures are essential for identifying sows and gilts that do not conceive soon after breeding. By using deep learning to automate this process, we can improve efficiency and effectiveness in these operations.

The implementation with more results and plots can be found on [GitHub](#).

2 Related Works

In the context of computer vision, various deep learning architectures have been developed to tackle complex image classification tasks. Among the most prominent are **Convolutional Neural Networks (CNNs)**, which have long been the standard due to their ability to capture spatial hierarchies in image data. More recently, **Vision Transformers (ViT)** have emerged as a powerful alternative, leveraging the self-attention mechanism to process images as sequences of patches [OCS⁺25]. Additionally, for tasks involving sequences or volumetric data, three-dimensional CNNs (3D CNNs) are often employed to capture temporal or depth-wise relationships.

Building on these foundational and contemporary approaches, this project compares the performance of a **ResNet18-based architecture** against both a ViT and a 3D CNN. The decision to select and adapt the ResNet18 model was informed by its proven effectiveness in similar image classification challenges. By evaluating its performance relative to these other state-of-the-art models, this study aims to demonstrate the practical utility of a hybrid CNN-based approach for the specific problem of embryo viability detection, where both spatial features and temporal changes (in the case of image sequences) are critical.

3 Proposed Approach

3.1 Data Exploration and Preprocessing

The dataset for this project comprised **1083 folders**, each containing **20 grayscale images** of pig embryos captured over time.

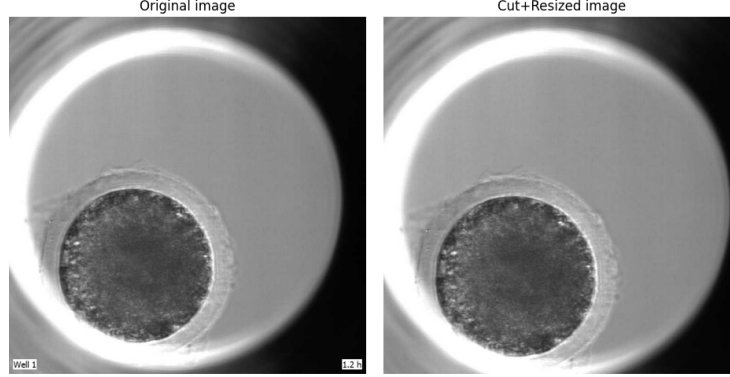


Figure 1: Cut and Resized image example

The primary challenge was the significant **imbalance** in the data, where non-viable embryos greatly outnumbered viable ones. This skew, if not addressed, could lead to a model that exhibits high overall accuracy but performs poorly on the minority class. To mitigate this and enable the model to generalize effectively, a comprehensive **data augmentation** strategy was employed after cutting and resizing the images to **256x256** dimension.

This involved applying a series of **transformations** to the training data, including random horizontal and vertical flips, random rotation, and color jitter. These augmentations effectively increased the size and diversity of the viable embryo class, creating a more balanced dataset for training and helping to prevent the model from overfitting. The augmentation has only be applied on Class 1 images, augmenting every folder 5 times.

The dataset included a **spreadsheet** containing the exact name of the image from which the embryo was found to be fertilized or not. Thanks to the implementation of a Python script, we were able to automate the labeling of individual images by changing their name with a final ₀ or ₁, based on the spreadsheet information.

Dataset has been splitted in three sets (train, validation and test) with a ratio of **70:15:15**. As data distribution is imbalanced, we made a balanced version only of the train split with a ratio of **1.5:1** with respect to Class 0 and Class 1. This balanced dataset has been only used to assess the baseline model. The main training has been performed on the full original dataset and the augmented version of it.

3.2 Model Architectures

We evaluated several architectures, including Vision Transformers (ViT) and 3D Convolutional Neural Networks (3D CNNs). Ultimately, the ResNet18-based architecture was selected for the final models due to its strong performance and efficiency. We developed two distinct models to address the different natures of our data:

- **ResNet18 + Binary Classification:** This model was designed for the classification of single images. The architecture leverages the ResNet18 backbone for feature extraction, followed by a binary classification layer to determine the viability of the embryo in the image. This approach is well-suited for a straightforward classification task where temporal data is not a factor.
- **ResNet18 + LSTM + Binary Classification** [VHD⁺20] [VCG⁺21]: For the time-lapse image sequences, a more sophisticated approach was required to capture the temporal dependencies between images. This model integrates a Long Short-Term Memory (LSTM) layer on top of the ResNet18 backbone. The ResNet18 layer first extracts features from each image in the sequence, and these features are then fed into the LSTM layer. The LSTM analyzes the sequence of features to understand the temporal progression of the embryo, providing a more robust basis for the final binary classification.

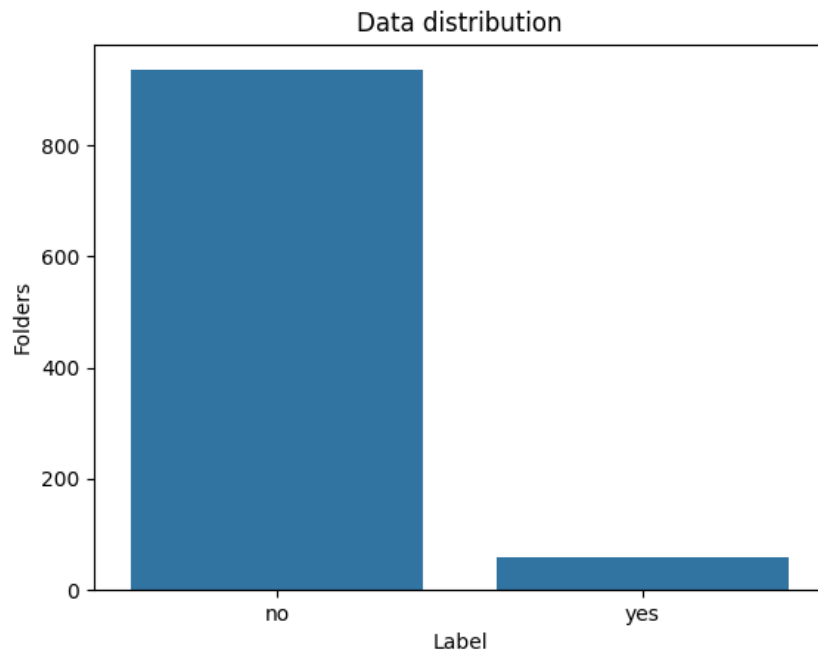


Figure 2: Label Distribution with respect to Folders

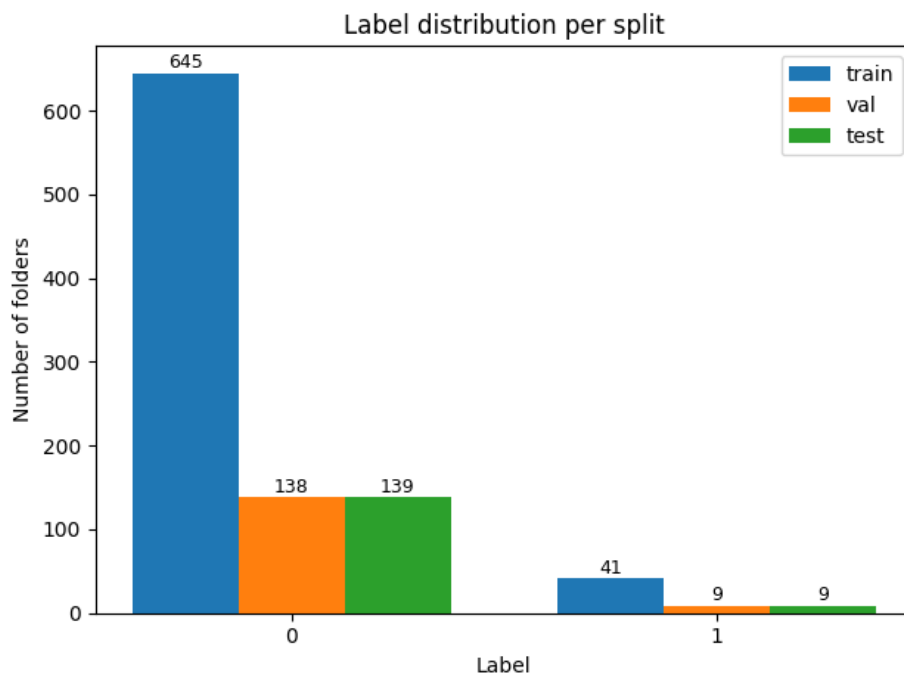


Figure 3: Label distribution per split

This dual-model approach allowed us to optimize our strategy for both single-image and sequence-based classification, ensuring that the model selected for each task was specifically tailored to its unique requirements.

4 Results

We evaluated the models based on accuracy, as well as precision, recall, and F1-score, which are particularly important for understanding performance on imbalanced datasets.

Initial evaluation of several baseline architectures, including a **3D Convolutional Neural Network (3D CNN)** for sequence data and a **Vision Transformer (ViT)**, demonstrated that they did not perform as effectively as the ResNet18-based models. Due to their superior performance, the **ResNet18** variants were chosen for the final, detailed analysis presented below.

4.1 ResNet18 + LSTM + Binary Classification

This model was designed to classify image sequences, leveraging an LSTM layer to capture temporal dependencies. Its performance on both the original and augmented datasets is summarized below.

Table 1: ResNet18 + LSTM + Binary Classification on the Original Dataset

	Precision	Recall	F1-Score	Support
Class 0	1.0000	0.9071	0.9513	140
Class 1	0.3810	1.0000	0.5517	8
Accuracy		0.9122		148
Macro AVG	0.6905	0.9536	0.7515	148
Weighted AVG	0.9665	0.9122	0.9297	148

On the **original dataset**, the model achieved an accuracy of 0.9122. While its performance on the majority "non-viable" class (Class 0) was excellent, the results for the minority "viable" class (Class 1) were mixed. The model had a perfect recall of 1.0000 for Class 1, meaning it correctly identified all viable embryos, but its precision was very low at 0.3810. This indicates that many of its "viable" predictions were actually incorrect, leading to a modest F1-score of 0.5517 for this class. The high overall accuracy was misleading as it was heavily skewed by the model's performance on the dominant Class 0.

Table 2: ResNet18 + LSTM + Binary Classification on the Augmented Dataset

	Precision	Recall	F1-Score	Support
Class 0	0.9655	1.0000	0.9825	140
Class 1	1.0000	0.3750	0.5455	8
Accuracy		0.9662		148
Macro AVG	0.9828	0.6875	0.7640	148
Weighted AVG	0.9674	0.9662	0.9588	148

With the **augmented dataset**, the overall accuracy improved significantly to 0.9662. The model's precision for the viable class (Class 1) became perfect at 1.0000, but this came at a cost to its recall, which dropped to 0.3750. This indicates that while every "viable" prediction was correct, the model missed a majority of the actual viable embryos. The F1-score remained relatively stable at 0.5455, highlighting the trade-off between precision and recall.

4.2 ResNet18 + Binary Classification

This model was used for single-image classification. Its performance metrics on both datasets are shown below.

Table 3: ResNet18 + Binary Classification on the Original Dataset

	Precision	Recall	F1-Score	Support
Class 0	0.9918	0.9935	0.9927	2930
Class 1	0.5128	0.4545	0.4819	44
Accuracy		0.9855		2974
Macro AVG	0.7523	0.7240	0.7373	2974
Weighted AVG	0.9847	0.9855	0.9851	2974

On the **original dataset**, the model achieved a high accuracy of 0.9855. Similar to the LSTM model, it performed well on the non-viable class (Class 0) but struggled with the minority viable class (Class 1), with a precision of 0.5128 and recall of 0.4545, resulting in a low F1-score of 0.4819.

Table 4: ResNet18 + Binary Classification on the Augmented Dataset

	Precision	Recall	F1-Score	Support
Class 0	0.9905	0.9922	0.9913	2930
Class 1	0.4103	0.3636	0.3855	44
Accuracy		0.9829		2974
Macro AVG	0.7004	0.6779	0.6884	2974
Weighted AVG	0.9801	0.9829	0.9823	2974

Interestingly, after training on the **augmented dataset**, the model’s performance on the minority class declined further. Both precision and recall for Class 1 decreased, resulting in an even lower F1-score of 0.3855. This finding highlights that data augmentation is not a one-size-fits-all solution; its effectiveness depends on the specific model and the nature of the data itself.

5 Conclusions

This project successfully developed and evaluated a computer vision model for the classification of pig embryo images. The core challenge of a severely imbalanced dataset was effectively mitigated through a robust data augmentation strategy. By applying techniques such as flips, rotations, and color jitter, we were able to create a more balanced training environment, which was essential for the model’s ability to generalize and avoid overfitting to the majority class.

Our analysis of the ResNet18-based models highlighted the critical importance of evaluating performance with metrics beyond simple accuracy. While both models achieved high overall accuracy, a deeper look at the precision, recall, and F1-score for the minority class of ”viable” embryos revealed more nuanced results.

- The **ResNet18 + LSTM model** demonstrated superior performance on the augmented dataset, achieving a high overall accuracy of 96.62%. For the minority class, it showed a significant improvement in precision to 1.0000, indicating that when it did predict an embryo was viable, it was always correct.
- The **ResNet18 + Binary Classification model** for single images, while also accurate, showed a decline in its ability to correctly classify the minority class on the augmented dataset, with both precision and recall decreasing. This suggests that the effectiveness of data augmentation is model-specific and not a universal solution for all architectures.

In conclusion, our findings demonstrate that computer vision is a powerful tool for embryo viability detection. The project successfully navigated the challenges of a difficult dataset through strategic data augmentation and model selection. Future work could focus on exploring additional architectures or different augmentation techniques to further enhance the models’ ability to detect the underrepresented ”viable” embryos.

References

- [OCS⁺25] Krittapat Onthuam, Norrawee Charnpinyo, Kornrapee Suthicharoenpanich, Suphaset Engphaiboon, Punnarai Siricharoen, Ronnapee Chaichaowarat, and Chanakarn Suebthawinkul. Combined input deep learning pipeline for embryo selection for in vitro fertilization using light microscopic images and additional features. *Journal of Imaging*, 11(1):13, 2025.
- [VCG⁺21] Gargee Vaidya, Shreya Chandrasekhar, Ruchi Gajjar, Nagendra Gajjar, Deven Patel, and Manish Banker. Time series prediction of viable embryo and automatic grading in ivf using deep learning. *The Open Biomedical Engineering Journal*, 15(1), 2021.
- [VHD⁺20] M VerMilyea, JMM Hall, SM Diakiw, A Johnston, T Nguyen, D Perugini, A Miller, A Picou, AP Murphy, and M Perugini. Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during ivf. *Human Reproduction*, 35(4):770–784, 2020.