Classification of Embryo Images for Viability Detection

Big Data Exam Project, AY 2024/2025

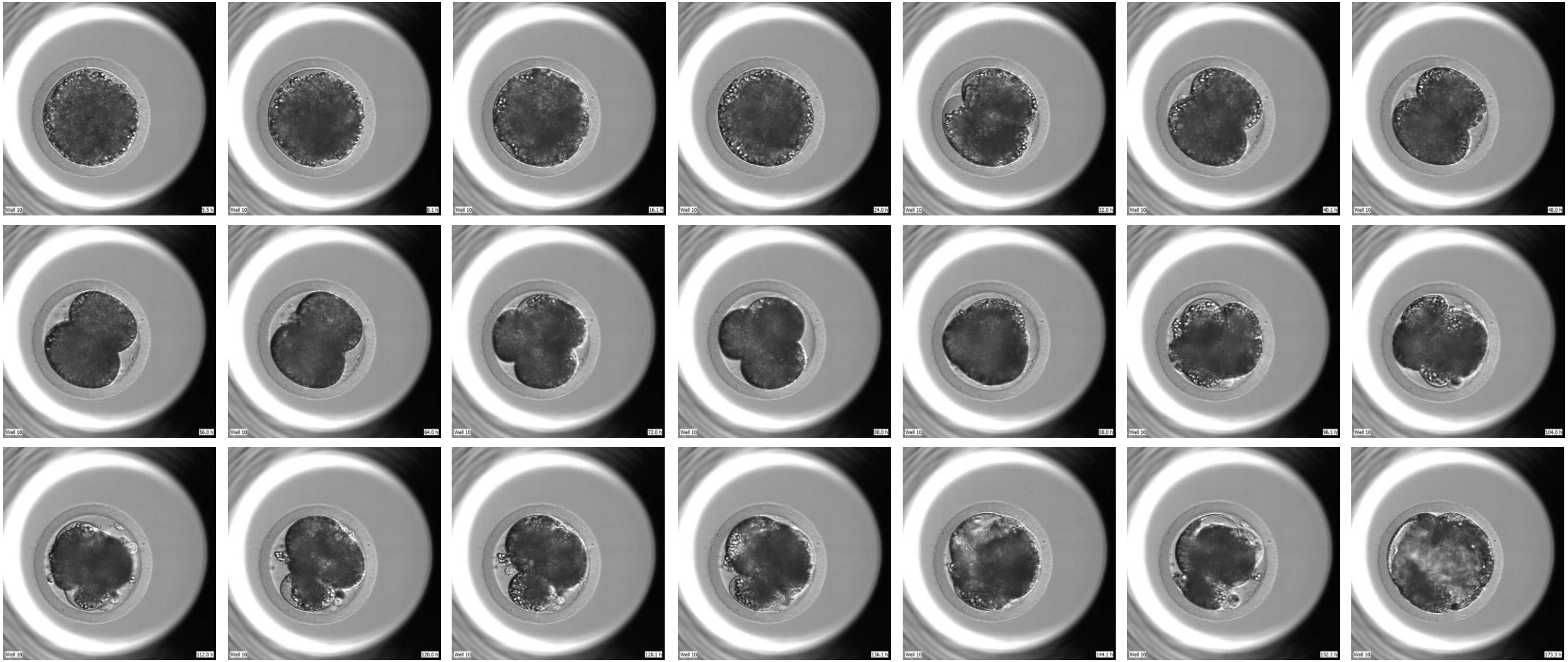
Introduction

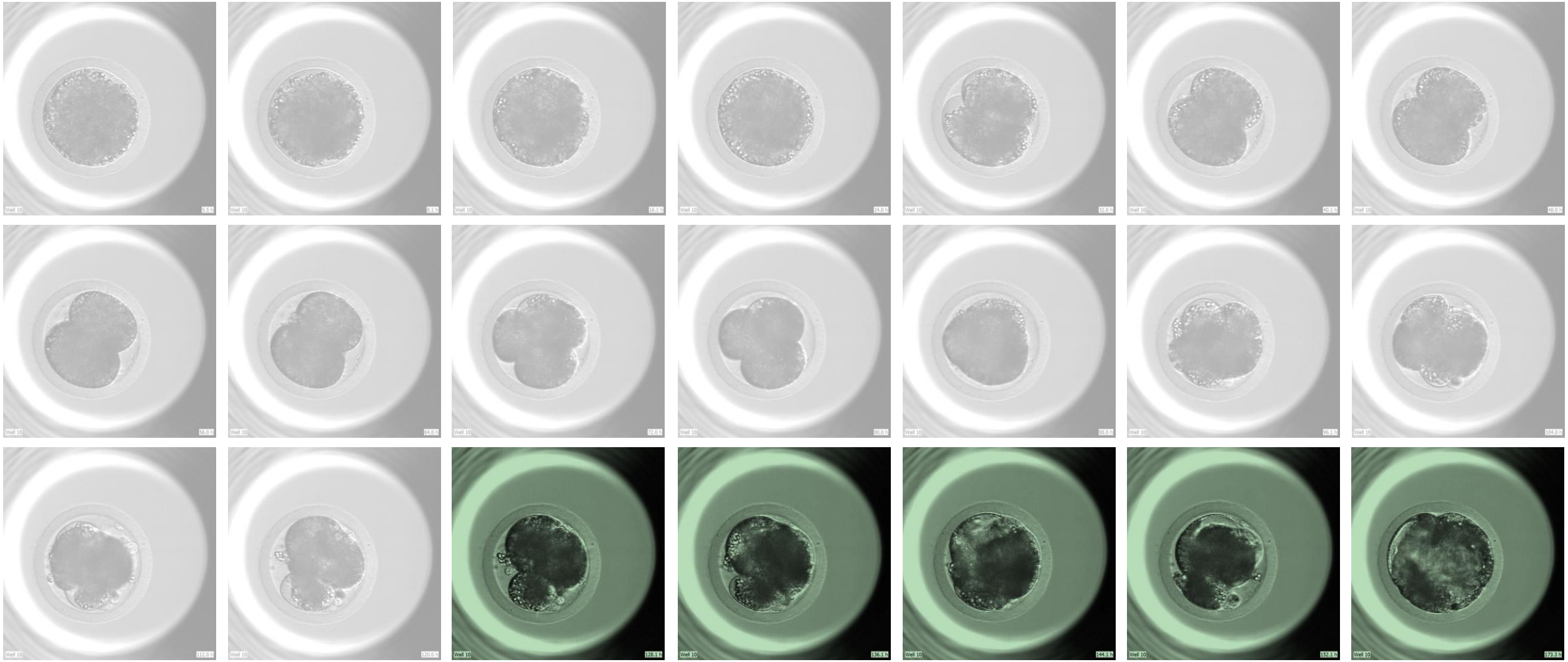
Objective

Train a computer vision model to classify images or sequences of images of pig embryos in order to assess their **viability**.

Why?

Effective pregnancy diagnostic procedures allows an **efficient reproductive management program** for swine operations identifying sows and gilts that do not conceice soon after breeding[1].





Problem Statement

Current challenge

To perform classification, **time-lapse imaging** is needed[2] and the rate at which viable embryos are detected is generally **very low** with respect to non-viable ones.

Goal

Classify which image sequences (or single images) will lead to **viable** embryos **or not**.

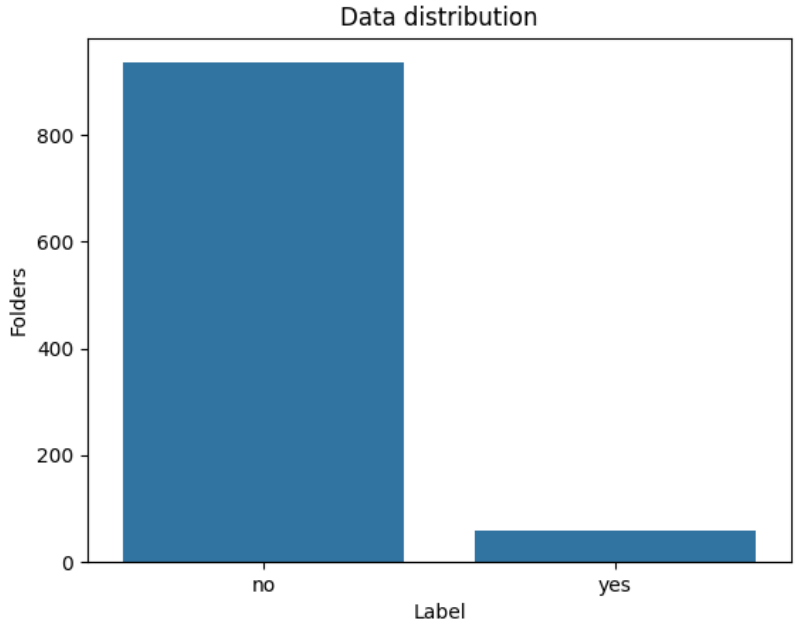
Data Collection

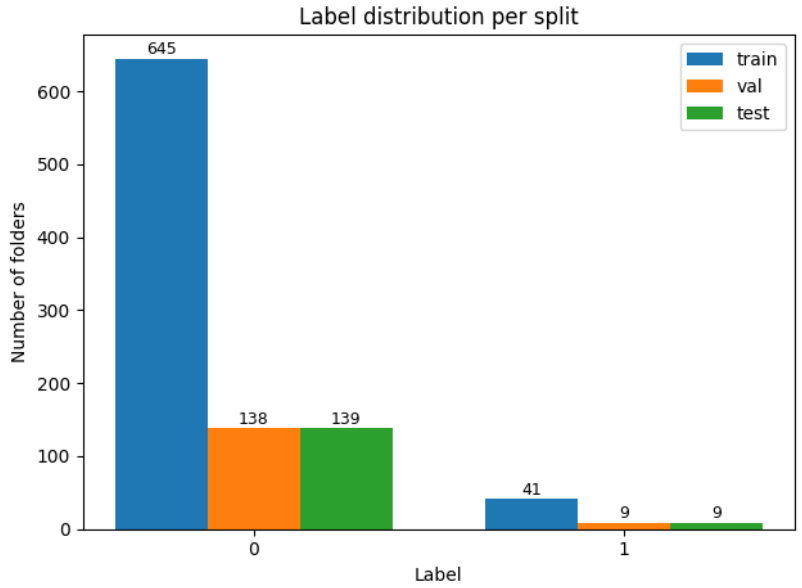
Data Source

Microscope images **sampled** and **labeled** by biologists, arranged in years, months and daily wells.

Volume

The dataset comprises **1083 folders** containing **20 grey-scale images** each over **3 years** of data collection.

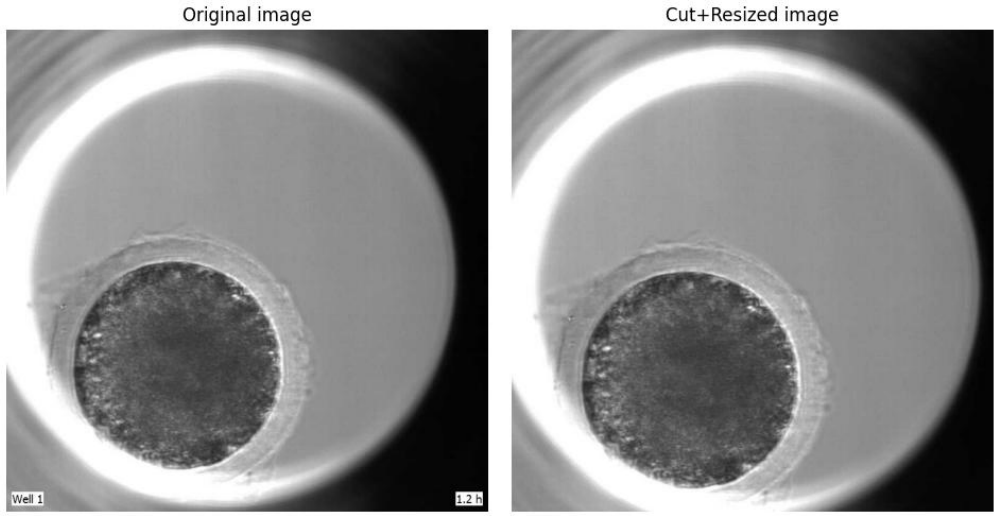




Data Processing

Cleaning

Images have been cut and resized to get **256x256** squared images before the split to remove well number and time from the image.



Data Distribution

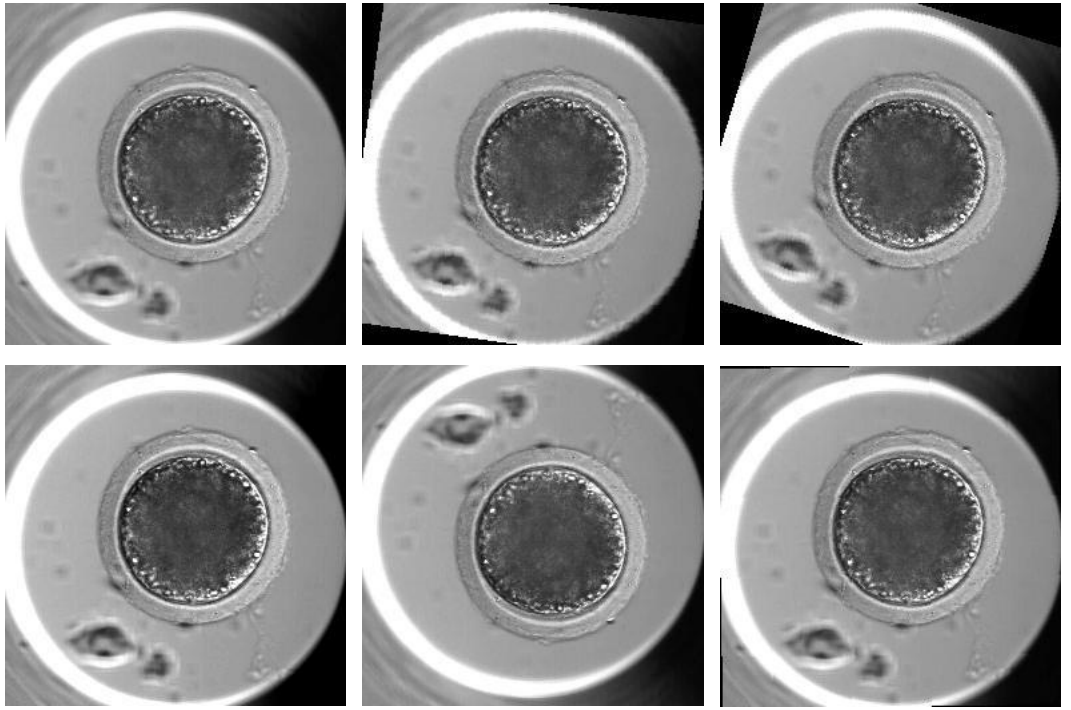
Challenges

The dataset is quite imbalanced and required some data augmentation procedures based on image transformations

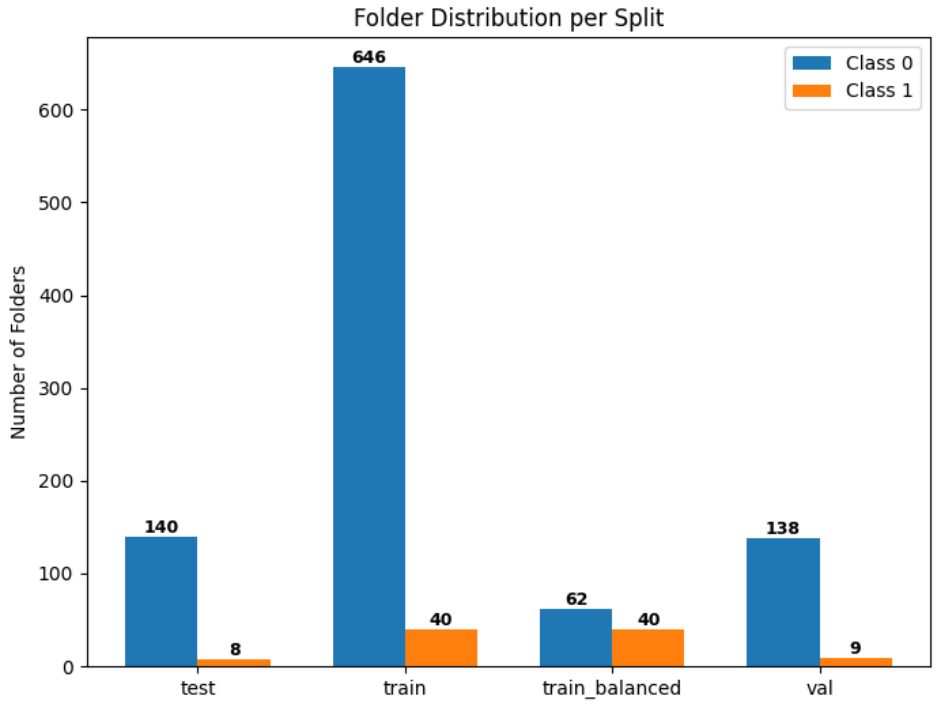
Transformations

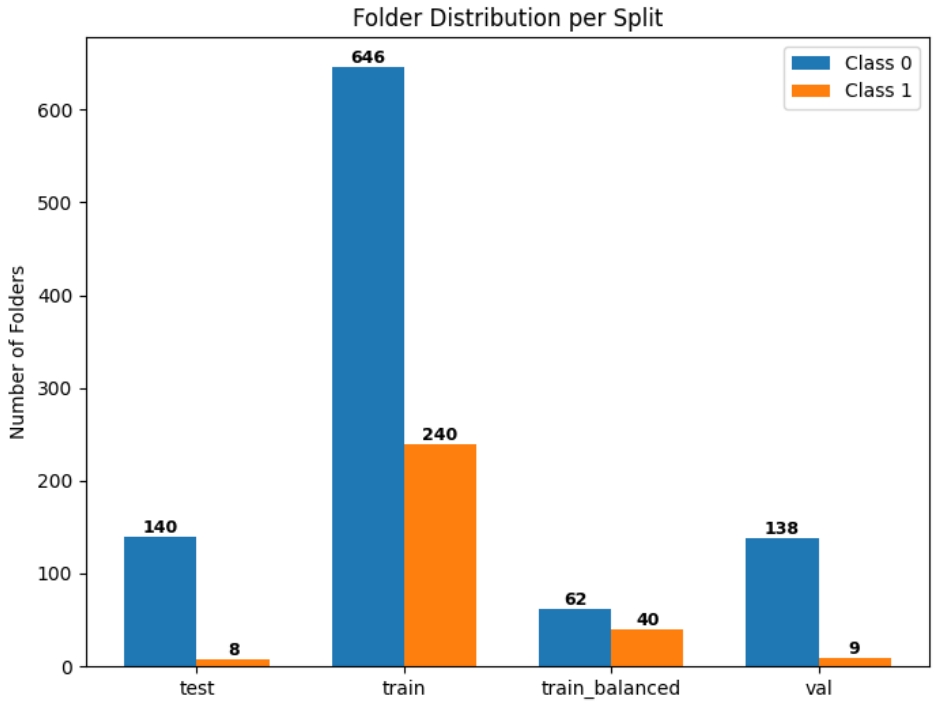
Data augmentation has been performed to reduce the imbalance of the dataset:

- Random Horizontal and Vertical Flip
- Random Rotation
- Color Jitter

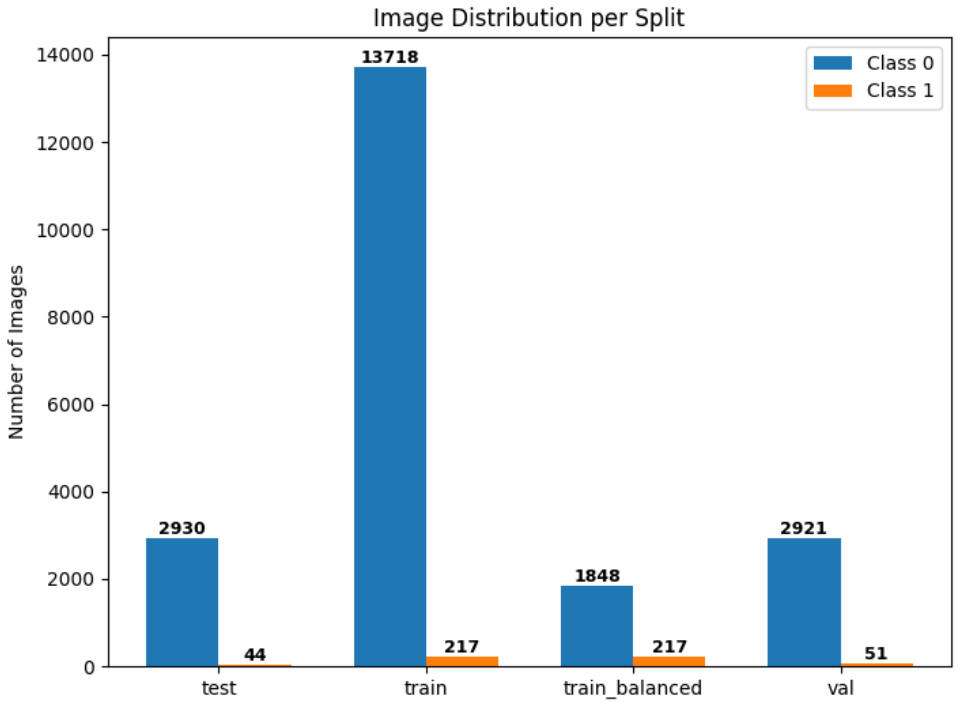


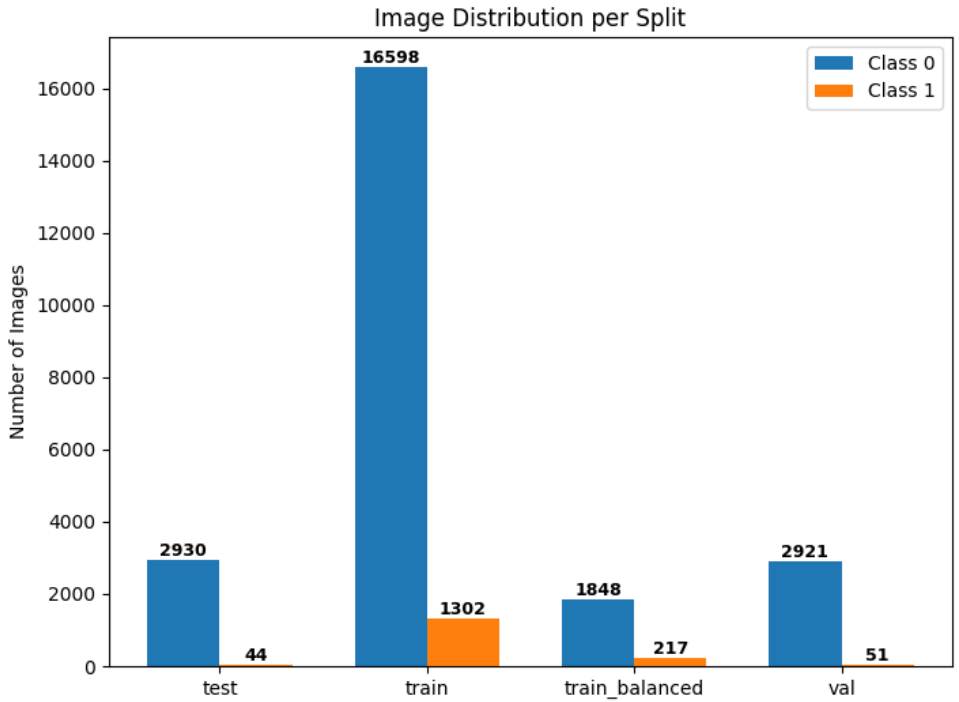
Before and After Data Augmentation





Before and After Data Augmentation





Baseline Selection

Baseline architectures

- ResNet18 + LSTM + Binary Classification[1,2,3]
- ViT + LSTM + Binary Classification[1]
- 3D CNN + Binary Classification

Why?

LSTM was crucial in order to catch the measure of time for the sequences of images while 3D CNN has been considered assuming each folder as a video made of 20 frames.

Classification Report on Balanced Dataset*

ResNet18 + LSTM + Bin. Class.

	Precision	Recall	F1-Score	Support
Class 0	1.0000	0.6857	0.8136	140
Class 1	0.1538	1.0000	0.2667	8
Accuracy			0.7027	148
Macro AVG	0.5769	0.8429	0.5401	148
Weighted AVG	0.9543	0.7027	0.7840	148

ViT + LSTM + Bin. Class.

	Precision	Recall	F1-Score	Support
Class 0	0.0000	0.0000	0.0000	140
Class 1	0.0541	1.0000	0.1602	8
Accuracy			0.0541	148
Macro AVG	0.0270	0.5000	0.0513	148
Weighted AVG	0.0029	0.0541	0.0055	148

*Trained on the balanced version of the original dataset that contains 62 Class 0 folders and 40 Class 1 folders

Classification Report on Balanced Dataset*

ResNet18 + LSTM + Bin. Class.

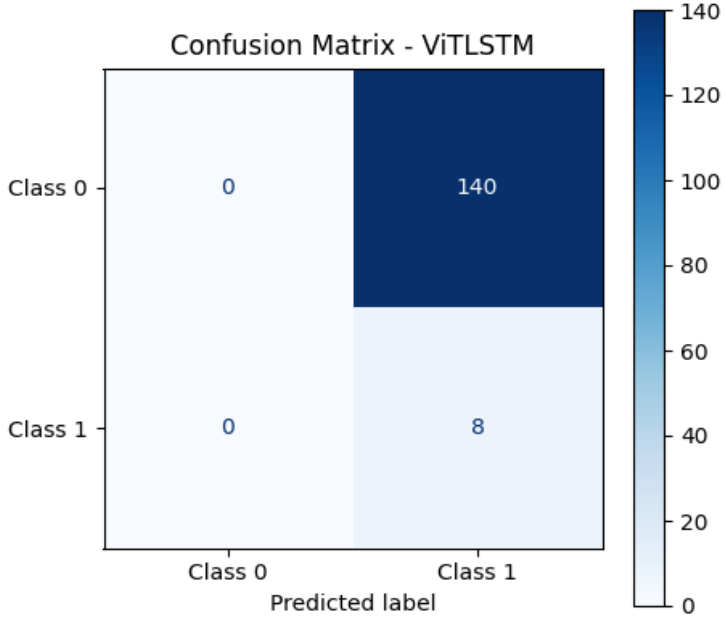
	Precision	Recall	F1-Score	Support
Class 0	1.0000	0.6857	0.8136	140
Class 1	0.1538	1.0000	0.2667	8
Accuracy			0.7027	148
Macro AVG	0.5769	0.8429	0.5401	148
Weighted AVG	0.9543	0.7027	0.7840	148

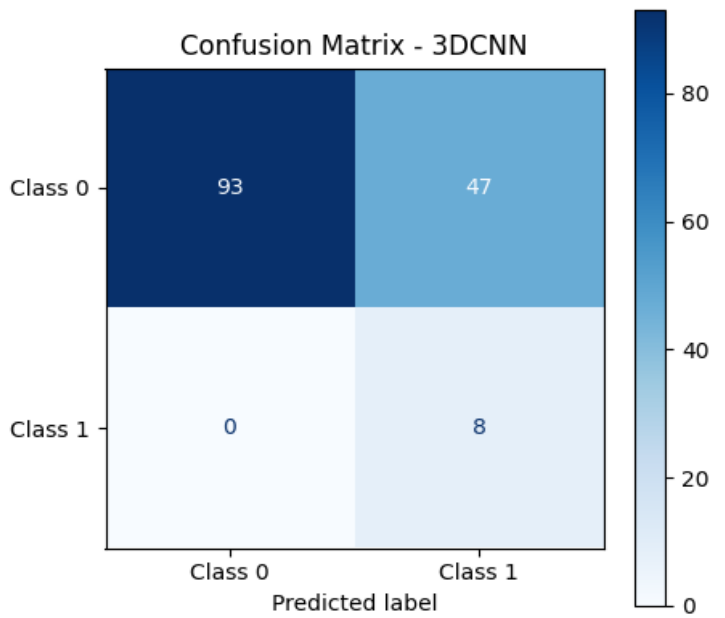
3D CNN

	Precision	Recall	F1-Score	Support
Class 0	0.9655	1.0000	0.9825	140
Class 1	1.0000	1.0000	0.2540	8
Accuracy			0.6824	148
Macro AVG	0.5727	0.8321	0.5261	148
Weighted AVG	0.9538	0.6824	0.7689	148

*Trained on a balanced version of the original dataset that contains 62 Class 0 folders and 40 Class 1 folders


Confusion Matrix - ResNet18LSTM
Class 0
Class 1
96
44
0
8
Class 0
Class 1
Predicted label
80
60
40
20
0


Confusion Matrix - ViTLSTM
Class 0
Class 1
0
140
0
8
Class 0
Class 1
Predicted label
140
120
100
80
60
40
20
0


Confusion Matrix - 3DCNN
Class 0
Class 1
93
47
0
8
Class 0
Class 1
Predicted label
80
60
40
20
0

ResNet18 + LSTM Classification Reports

ResNet18 on balanced dataset*

	Precision	Recall	F1-Score	Support
Class 0	1.0000	0.6857	0.8136	140
Class 1	0.1538	1.0000	0.2667	8
Accuracy			0.7027	148
Macro AVG	0.5769	0.8429	0.5401	148
Weighted AVG	0.9543	0.7027	0.7840	148

ResNet18 on augmented balanced dataset*

	Precision	Recall	F1-Score	Support
Class 0	1.0000	0.8786	0.9354	140
Class 1	0.3200	1.0000	0.4848	8
Accuracy			0.8851	148
Macro AVG	0.6600	0.9393	0.7101	148
Weighted AVG	0.9632	0.8851	0.9110	148

*Trained on a balanced version of the original dataset that contains 62 Class 0 folders and 40 Class 1 folders

Model Selection

Selected architectures

- ResNet18 + LSTM + Binary Classification (for image sequences)
- ResNet18 + Binary Classification (for single images)

Why?

Architectured with ResNet18 performed better with respect to the others, so we propose two version: the former takes into account time while the latter does not.

Model Training

Training data

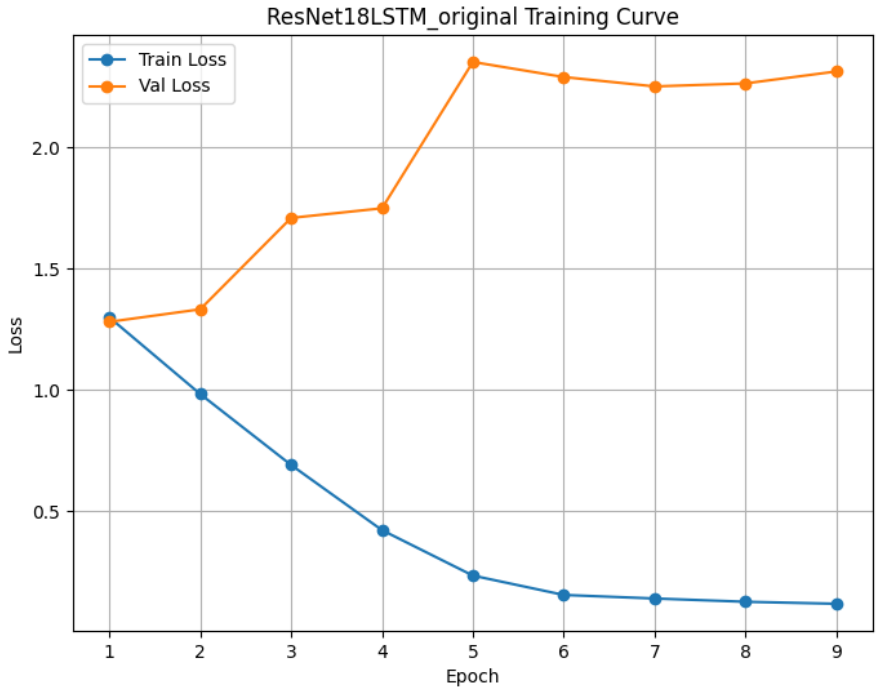
- **70%** of the dataset for **training (645 + 41 folders)**
- **15%** of the dataset for **validation (138 + 9 folders)**
- **15%** of the dataset for **testing (139 + 9 folders)**

Parameters

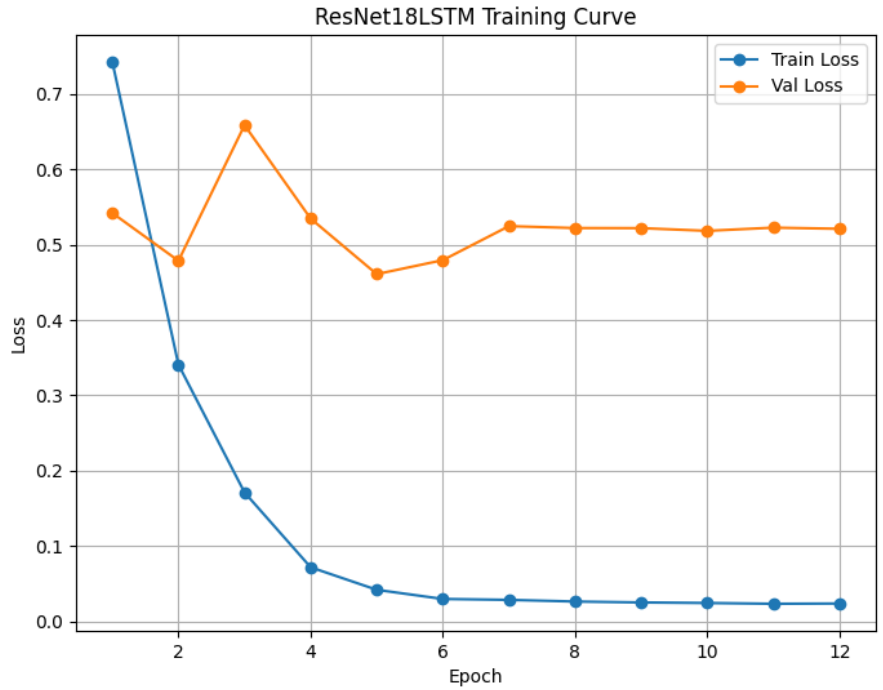
- **100 epochs** with early stopping (patience set to 5)
- **64 batch size** (max sequence length of 20 for LSTM architecture only)
- **AdamW** with learning rate of 10^{-4} and weight decay of 10^{-4}

ResNet18 + LSTM + Bin. Class. Loss

On the original dataset



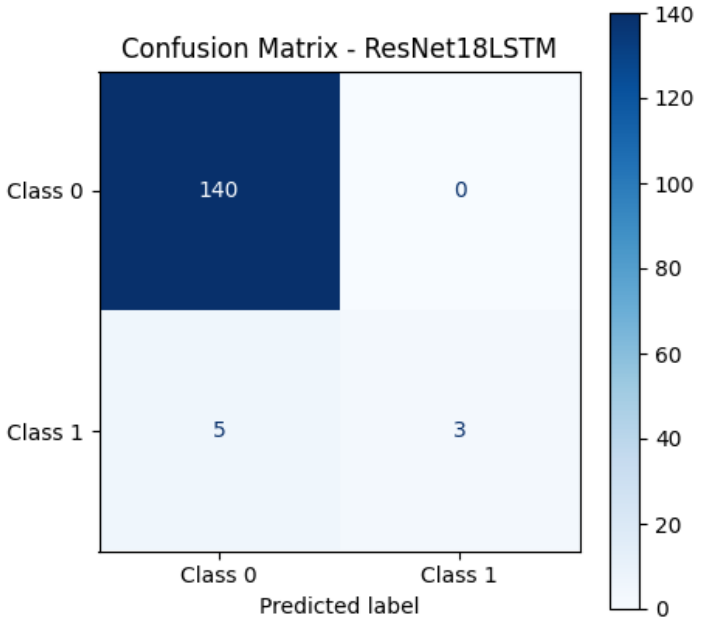
On the augmented dataset



ResNet18 + LSTM + Bin. Class. Confusion Matrix

On the original dataset

On the augmented dataset



ResNet18 + LSTM + Bin. Class. Report

On the original dataset

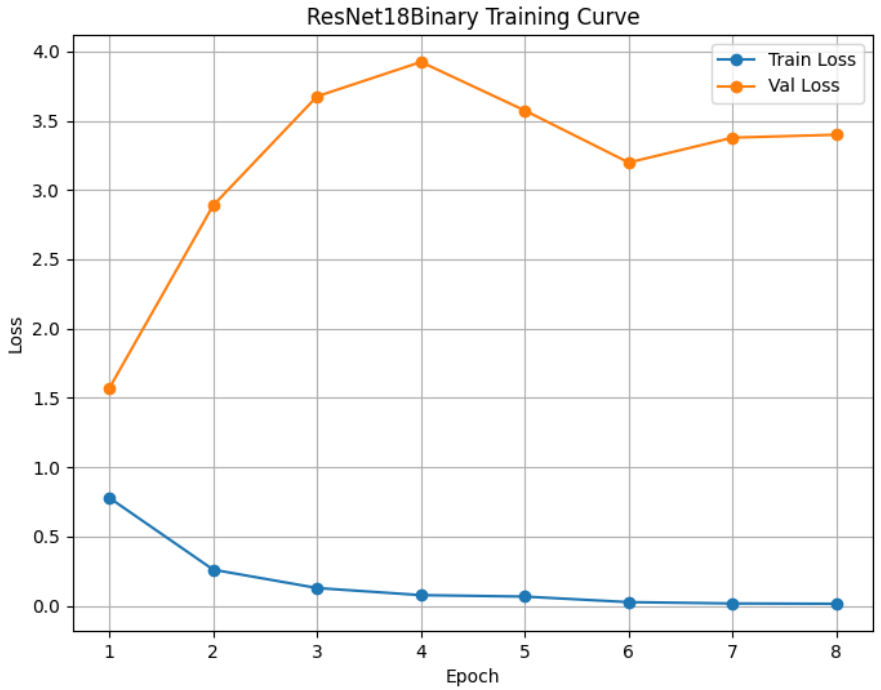
	Precision	Recall	F1-Score	Support
Class 0	1.0000	0.9071	0.9513	140
Class 1	0.3810	1.0000	0.5517	8
Accuracy			0.9122	148
Macro AVG	0.6905	0.9536	0.7515	148
Weighted AVG	0.9665	0.9122	0.9297	148

On the augmented dataset

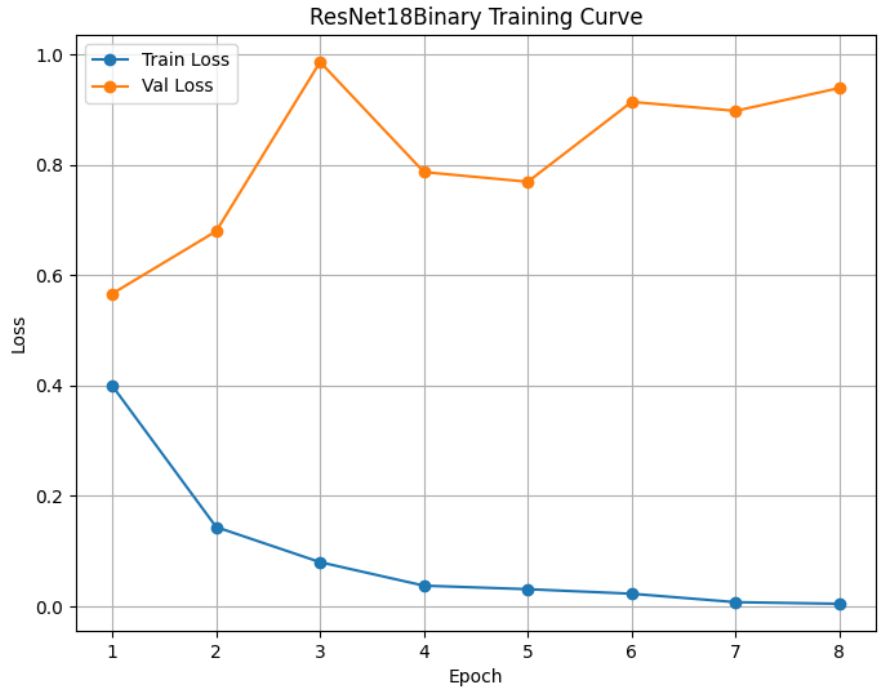
	Precision	Recall	F1-Score	Support
Class 0	0.9655	1.0000	0.9825	140
Class 1	1.0000	0.3750	0.5455	8
Accuracy			0.9662	148
Macro AVG	0.9828	0.6875	0.7640	148
Weighted AVG	0.9674	0.9662	0.9588	148

ResNet18 + Bin. Class. Loss

On the original dataset

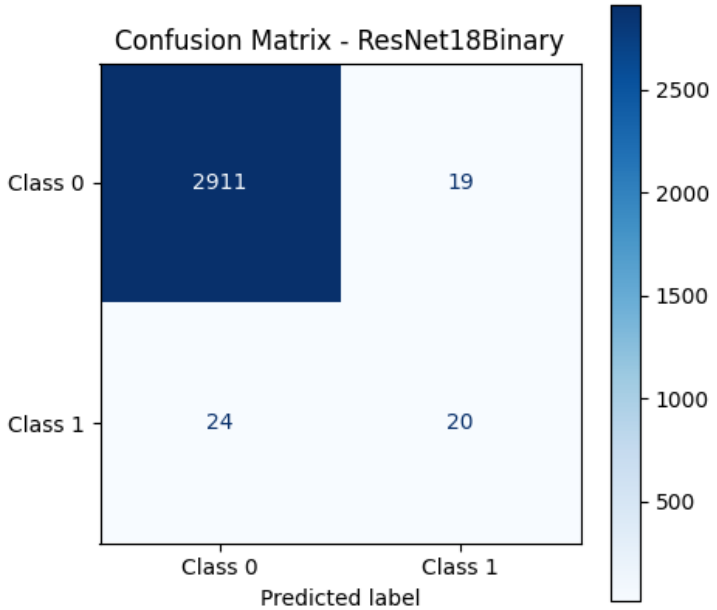


On the augmented dataset

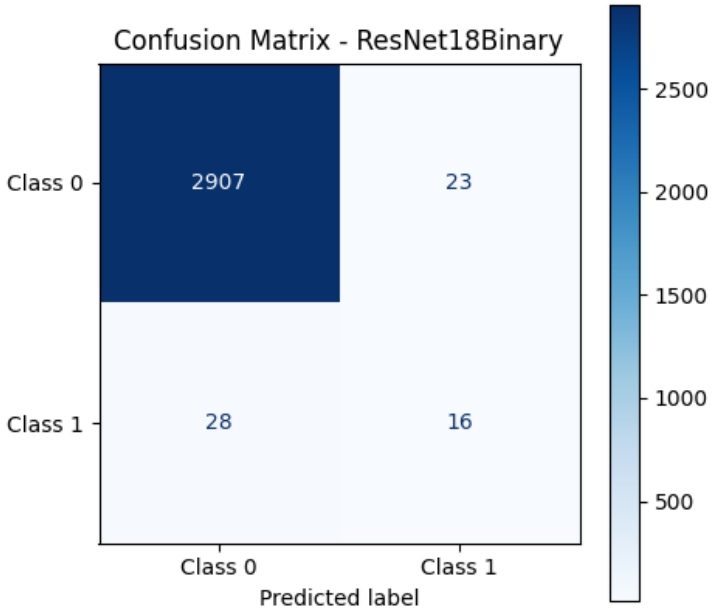


ResNet18 + Bin. Class. Confusion Matrix

On the original dataset



On the augmented dataset



ResNet18 + Bin. Class. Report

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 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

Classification Report on Augmented Dataset

ResNet18 + LSTM + Bin. Class

	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

ResNet18 + Bin. Class

	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

Conclusions

Data augmentation proved to be useful on the original dataset

- It **improved** the accuracy of ResNet18 + LSTM architecture (from 91.22% to 96.62%), but also other metrics
- In the other hand, it **did not improve** ResNet18 accuracy on single images (from 98.55% to 98.28%)

Both models achieved high **overall accuracy but** struggle with **low precision, recall, and F1-score for the minority class**.

References

1. Onthuam, Krittapat, et al. "Combined Input Deep Learning Pipeline for Embryo Selection for In Vitro Fertilization Using Light Microscopic Images and Additional Features." *Journal of Imaging* 11.1 (2025): 13.
2. 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." *Human Reproduction* 35.4 (2020): 770-784.
3. Vaidya, Gargee, et al. "Time series prediction of viable embryo and automatic grading in IVF using deep learning." *The Open Biomedical Engineering Journal* 15.1 (2021).