



**HAL**  
open science

## Early prediction of the transferability of bovine embryos from videomicroscopy

Yasmine Hachani, Patrick Bouthemy, Sylvie Ruffini, Ludivine Laffont, Elisa Fromont, Alline de Paula Reis

► **To cite this version:**

Yasmine Hachani, Patrick Bouthemy, Sylvie Ruffini, Ludivine Laffont, Elisa Fromont, et al.. Early prediction of the transferability of bovine embryos from videomicroscopy. IABM 2024 - Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale, Mar 2024, Grenoble, France. pp.1-1, 2024. hal-04880124

**HAL Id: hal-04880124**

**<https://hal.science/hal-04880124v1>**

Submitted on 10 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# EARLY PREDICTION OF THE TRANSFERABILITY OF BOVINE EMBRYOS FROM VIDEOMICROSCOPY



Y. Hachani<sup>1</sup>, P. Bouthemy<sup>1</sup>, S. Ruffini<sup>4</sup>, L. Laffont<sup>4</sup>, E. Fromont<sup>1,2</sup>, A. De Paula Reis<sup>3,4</sup>

<sup>1</sup>Inria Rennes, <sup>2</sup>Univ Rennes, IUF, IRISA, <sup>3</sup>Ecole Nationale Vétérinaire d'Alfort, <sup>4</sup>INRAE, BREED



## OVERVIEW

We aim to predict **early and automatically** the transferability of embryos for biological issues and cattle breeding challenges, taking as input **2D time-lapse videos** up to the fourth embryonic development day. We propose:

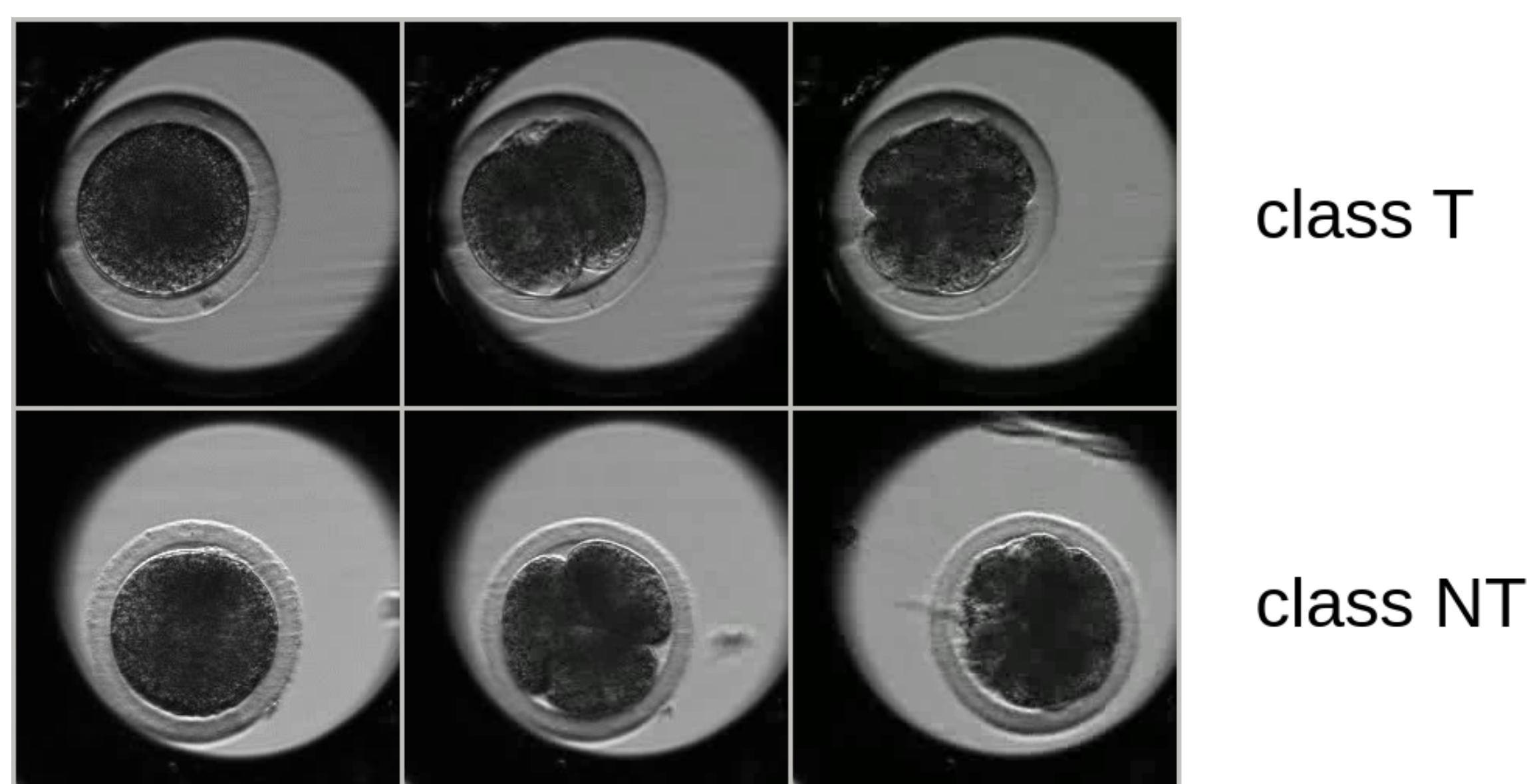
- a formulation as a **supervised classification** with two classes **Transferable (T)** and **Non Transferable (NT)**,
- a **3D-CNN** with **three pathways**, multi-scale in time and able to handle appearance and motion in different ways,
- the use of the **focal loss** for training.

Our model SFR compares favorably to other methods and demonstrates its accuracy and efficiency for our challenging biological task.

## DATA AND METHODOLOGICAL CHALLENGES

This problem is challenging for multiple reasons:

- Complex morphological and temporal processes
- Noisy, poorly contrasted videos with transparency effects and intertwined appearance and motion
- High intra-class variability and low inter-class distance
- Small set of 900 annotated videos
- Light annotation with only the class label per video



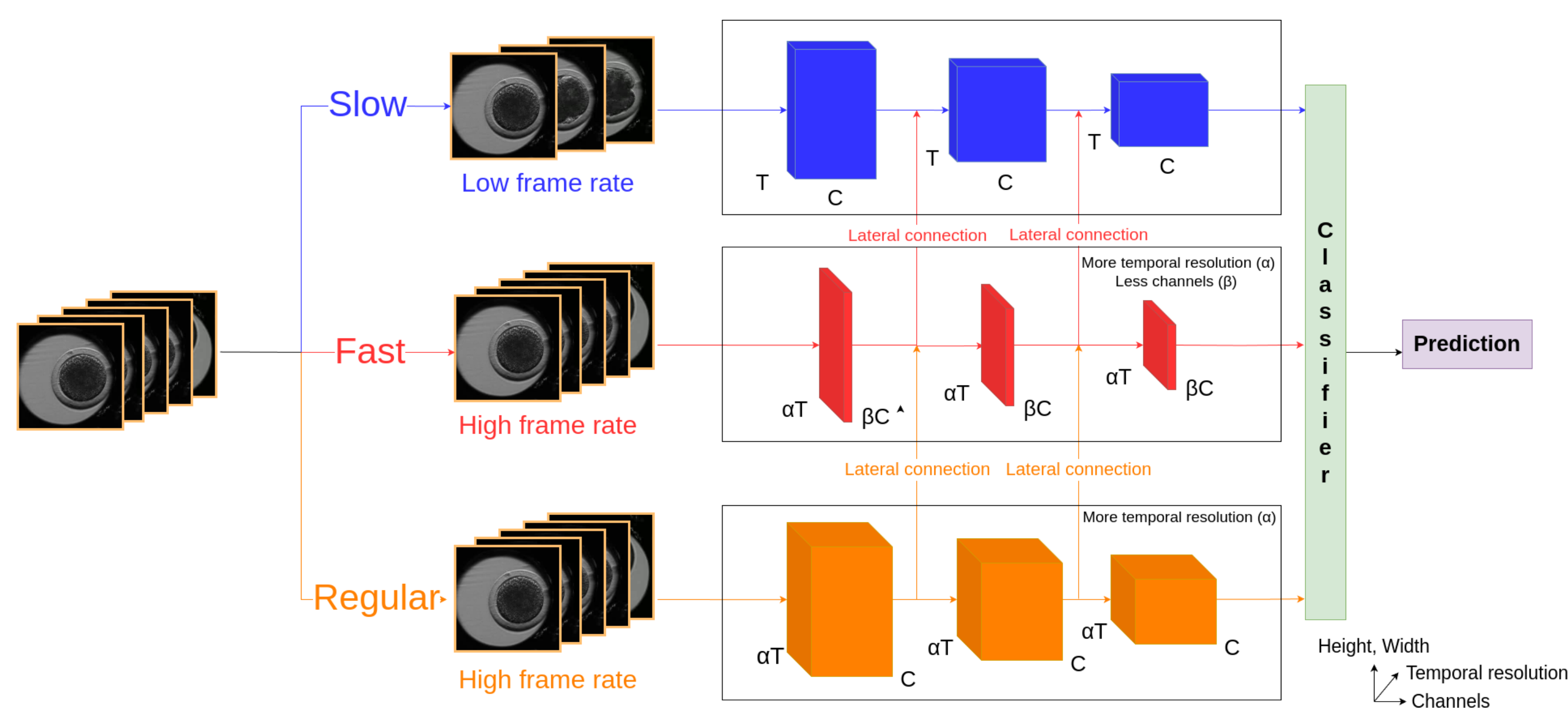
**Fig. 1:** Two sample videos of IVF bovine embryos, with three images taken at distant time instants. The bovine embryo (dark grey), surrounded by the zona pellucida, is located in a micro-well (light grey) within the Petri dish (black).

## OUR NETWORK CALLED SFR

Inspired by SlowFast [1], we have designed a 3D-CNN with three pathways handling appearance and motion in complementary ways:

- *Slow* pathway mainly dedicated to capture spatial features,
- *Fast* pathway mainly devoted to temporal features,
- *Regular* pathway combining appearance and motion.

All options were experimentally validated in an ablation study.



**Fig. 2:** Our model called SFR involves three pathways, *Slow*, *Fast*, and *Regular*, all implemented with 3D ResNet18 of different configurations. Input video is given at a  $\alpha$  times lower temporal rate for the *Slow* pathway. The *Fast* pathway has only a fraction  $\beta$  of the channels. The three outputs are combined before providing the prediction. Directed lateral connections are included between the pathways.

## FOCAL LOSS

We use the focal loss [2] instead of the cross entropy loss

- to address imbalance between the two classes,
- to focus the training of the model on hard examples,
- to improve performance (validated by the ablation study).

The focal loss writes:

$$\mathcal{L}_f(v, y) = - \sum_{c=1}^2 \eta_c (1 - \hat{p}(y_c|v))^{\gamma} p(y_c|v) \log \hat{p}(y_c|v),$$

with  $v$  the input video,  $y_c$  label of one of the two classes,  $\hat{p}(y_c|v)$  the predicted probability of having class  $c$  given video  $v$ ,  $p(y_c|v)$  the true one (equal to 1 here due to supervised learning),  $\eta_c$  the weight for class  $c$ , and  $\gamma$  the focusing parameter.

## IMPLEMENTATION DETAILS

- Use AdamW optimizer with a learning rate of  $10^{-4}$  and a cyclic learning rate scheduler,
- Set loss parameters to  $\gamma = 2$ ,  $\eta_1 = 1.25$ , and  $\eta_2 = 0.833$ ,
- Apply stochastic weight averaging technique [3].

## EXPERIMENTAL RESULTS

Model	Acc	$P_T$	$R_T$	$P_{NT}$	$R_{NT}$
2D-Resnet18+GRU	67.6±4.1	54.6±5.6	48.4±19.9	74.2±5.9	78.3±8.6
3D-Resnet18(FL)	74.6±4.0	64.1±6.5	<b>67.4±12.4</b>	<b>82.0±4.6</b>	78.6±7.8
SlowFast(FL)	73.1±2.6	63.9±4.0	55.2±9.3	77.3±3.2	82.8±3.4
Ours - SFR(FL)	<b>75.6±1.5</b>	<b>66.8±3.3</b>	62.5±3.4	80.1±1.1	<b>82.8±2.9</b>

**Table 1:** Accuracy, Precision and Recall, for the two classes  $T$  and  $NT$ , obtained with respectively 2D-Resnet18+GRU, 3D-Resnet18, SlowFast [1], and our SFR model, all the models with the focal loss ( $\gamma = 2$ ).

- Our SFR method provides the overall best performance,
- Biologists prioritize stability and our model is the most stable one,
- Key metrics are  $P_T$  and  $R_{NT}$  regarding the application needs, and our method provides the best scores for them.

## ACKNOWLEDGEMENTS

The research to produce the original data was funded by CRB-Anim and APIS-GENE.

## REFERENCES

- [1] C. Feichtenhofer, H. Fan, J. Malik, and K. He. SlowFast networks for video recognition. ICCV 2019.
- [2] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. ICCV 2017.
- [3] P. Izmailov, D. Podoprikin, T. Garipov, D. Vetrov, and A. Gordon Wilson. Averaging weights leads to wider optima and better generalization. UAI 2018.