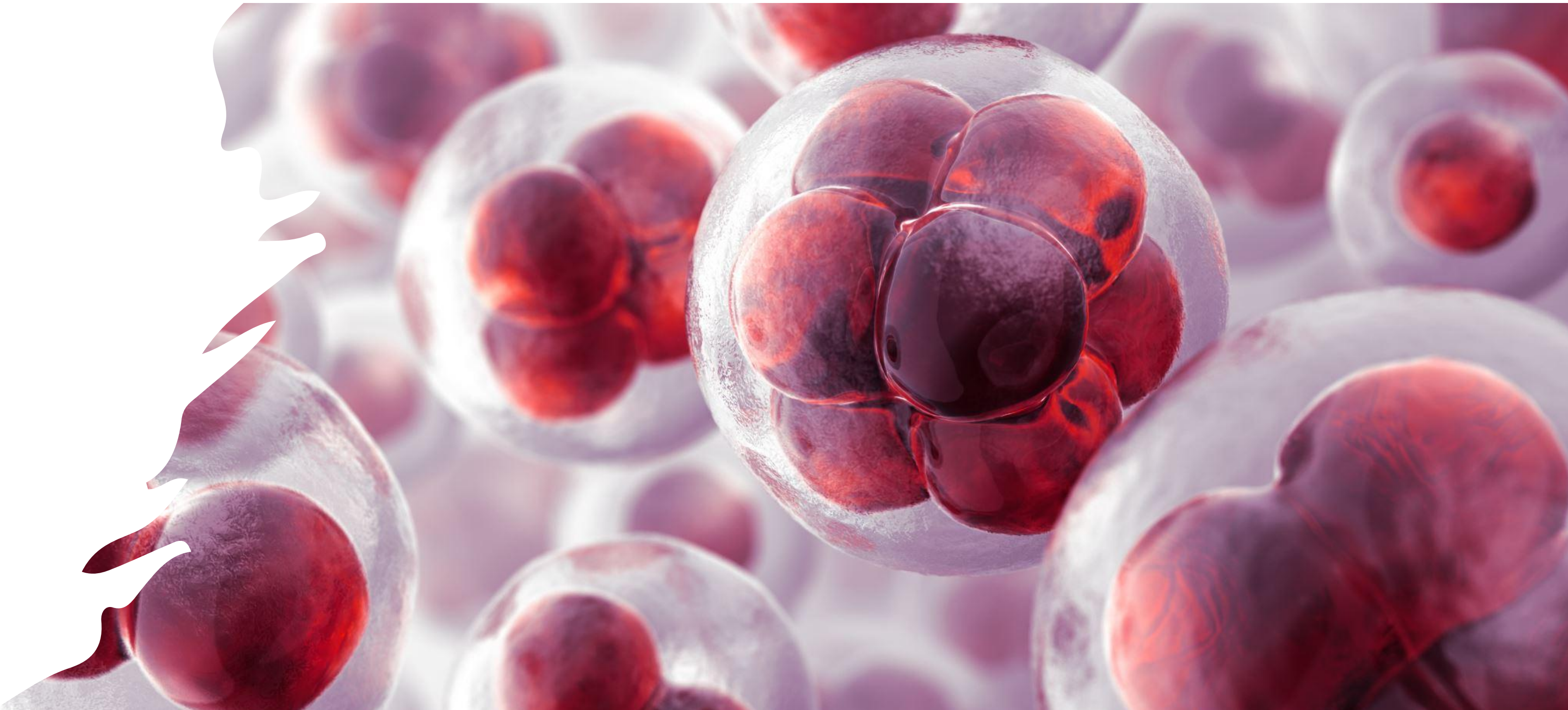
Invasive Ductal Carcinoma Classification



Team

- **Alessandro Tiveron**
  - Implementation of original project with fasAI library
- **Francesco Ferronato**
- **Mattia Tortelli**
- **Francesca Virgolini**
- **Niccolò Brusadin**

Index

- Description of the Original Project
- Limitations of the Original Project
- Objective
- Method
- Results

Data Overview and Task

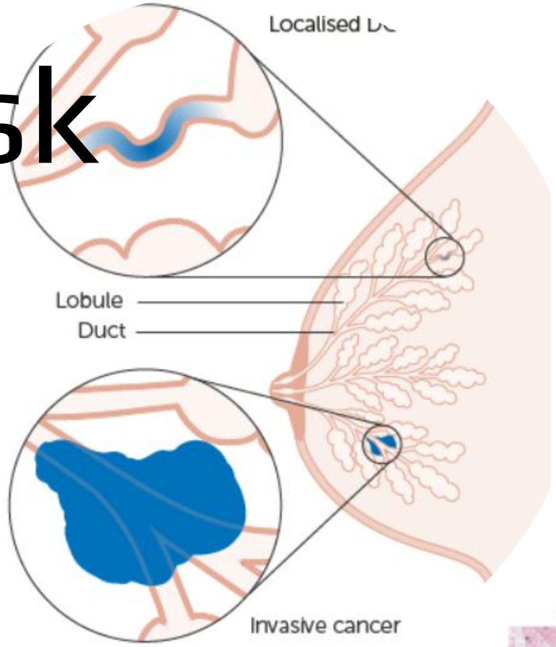
Original Dataset: 162 whole-mount slide images of Breast Cancer specimens

Patch Extraction: 277,524 patches (50 x 50 pixels)

Class Distribution:

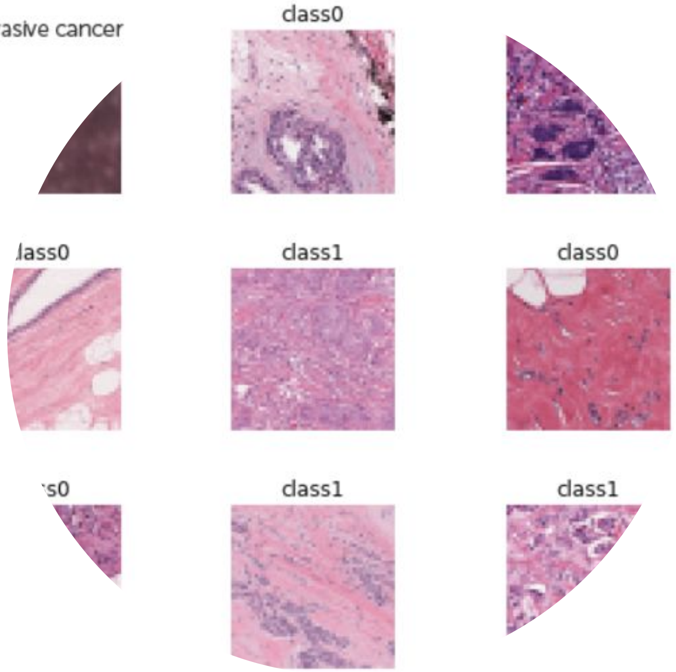
198,738 IDC negative

78,786 IDC positive



Filename Format: u_xX_yY_classC.png

- u: Patient ID (e.g., 10253_idx5)
- X: x-coordinate of patch
- Y: y-coordinate of patch
- C: Class (0 for non-IDC, 1 for IDC)

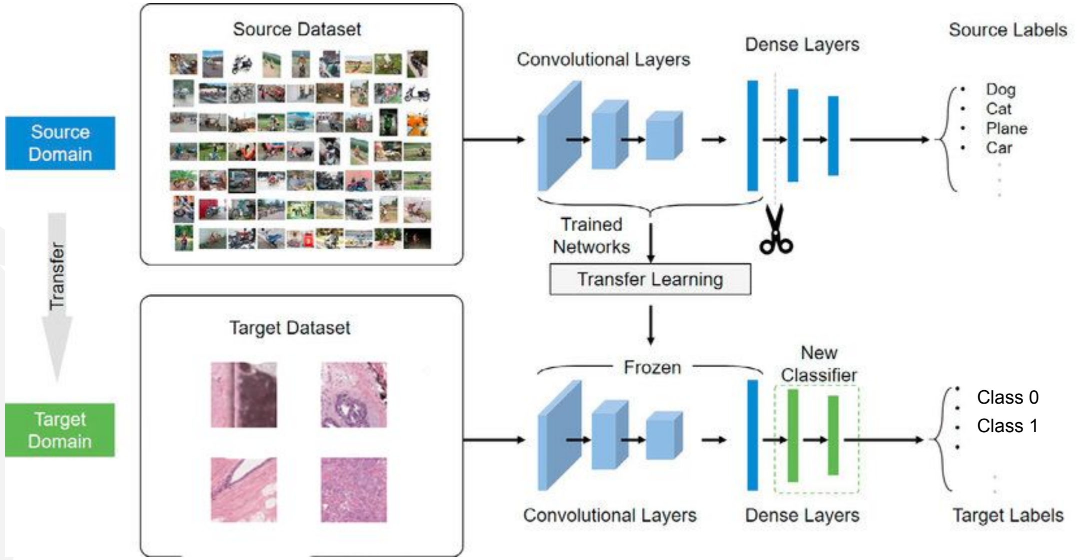


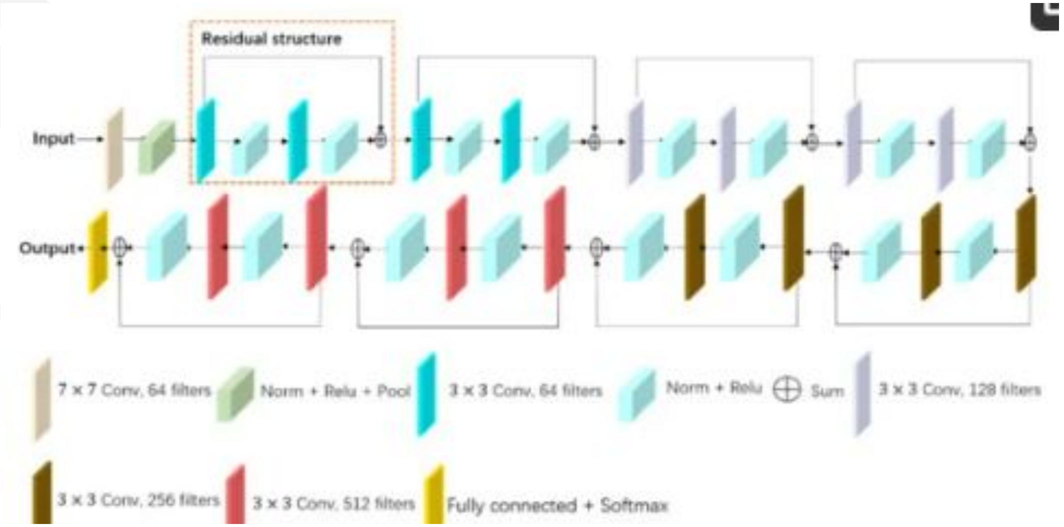
Objective: Classify patches into IDC negative (0) and IDC positive (1) using fastAI and CNN

Challenge: Utilize the dataset to develop a robust model for accurate Breast Cancer classification

Description of the Original Project

Methodology:	•FastAI implementation of Transfer Learning Approach
Data Exploration:	•Organized Kaggle dataset into patient folders •Each patient folder contains 2 subfolders 'class 0' and 'class 1'
Data Loading:	•ImageDataBunch setup with validation, augmentation, and normalization
Transfer Learning:	•Adopted pre-trained ResNet18 on ImageNet •Focus on fine-tuning last layer
Model Implementation:	•Implemented with cnn_learner() using pre-trained ResNet18
Learning Rate Optimization:	•Optimized using lr_find(), one-cycle training, and refined learning rates





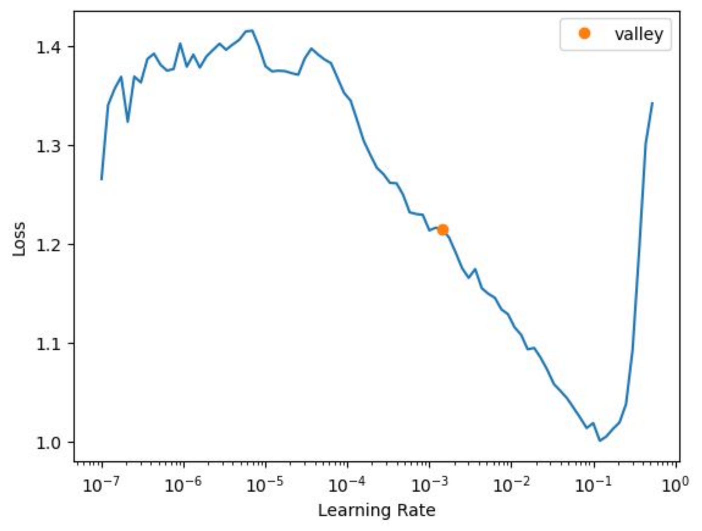
Results

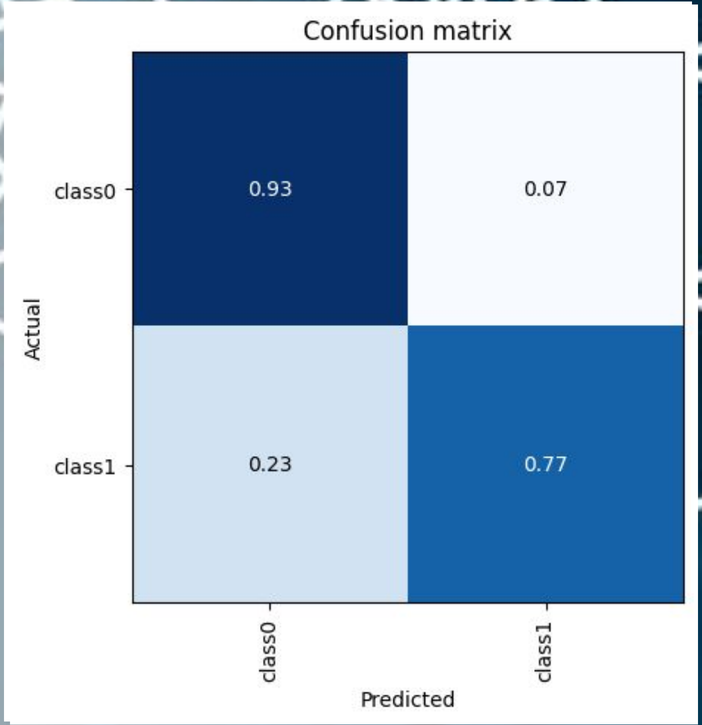
Performance of model trained with fit_one_cycle policy (for 1 epoch??) (Ale?)

epoch	train_loss	valid_loss	accuracy	time
0	0.311559	0.409009	0.879324	08:59

Performance improved after unfreezing and optimizing hyperparameters

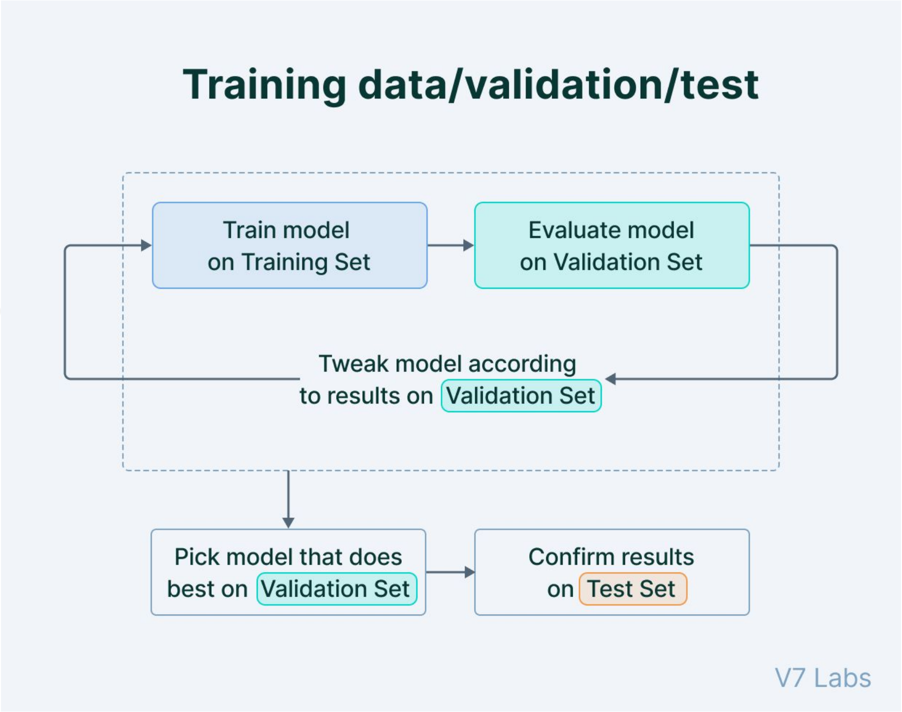
epoch	train_loss	valid_loss	accuracy	time
0	0.282195	0.790487	0.889035	09:19





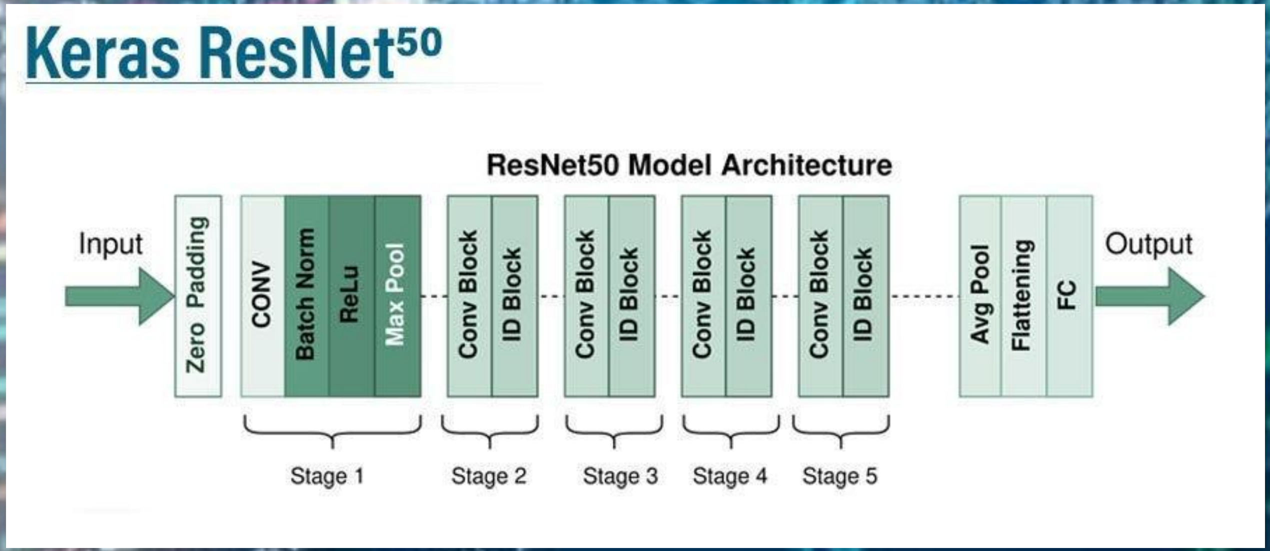
Limitations of the Original Project

- Outdated fastAI library version
- Data preparation
- Lack of test set
- Unbalanced classes



Objective

- Understanding the original project
- Transition from fastAI to Keras
- Implement the latest fastAI libraries
- Enhance robustness with a test set, balance classes
- Exploring ResNet50's potential



Methods: Pre-training

Library Utilization:

A Keras user-friendly implementation (to control data augmentation and preprocessing)

Data Preprocessing:

Organizing histopathology images into 70% training and 15% validation and 15% test set

Class Imbalance Handling:

Weights assigned to prioritize correct classification of class 1

Augmentation and Preprocessing:

With RandomFlip(), RandomRotation() and RandomZoom()

Methods: Model Training optimization

1. Finding Optimal Learning Rate

Identify the best learning rate for model training.

2. Unfreezing Model Parameters

Unlock all hyperparameters for comprehensive training

3. Re-evaluating Learning Rate Post-Unfreezing

Reassess the optimal learning rate after unfreezing parameters

4. Fine-Tuning

Define model convergence with unfrozen parameters

Methods: Models

Transfer Learning with EfficientNetV2S:

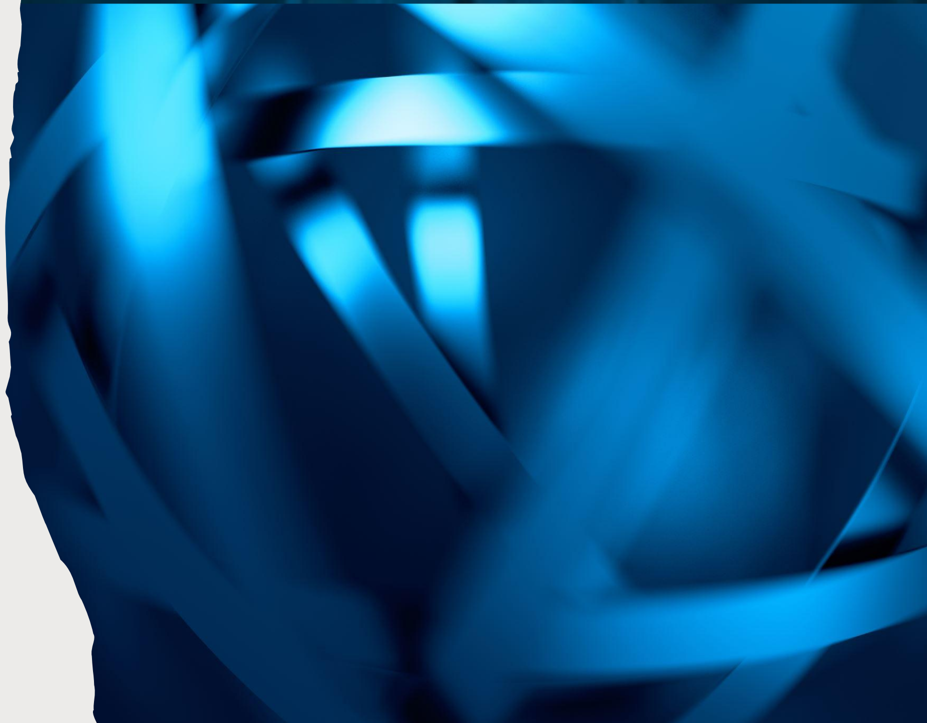
EfficientNetV2S, pretrained on ImageNet, as first model implementation

Transfer Learning with ResNet50:

Exploration of ResNet50 for binary classification as an alternative approach

Performance Evaluation:

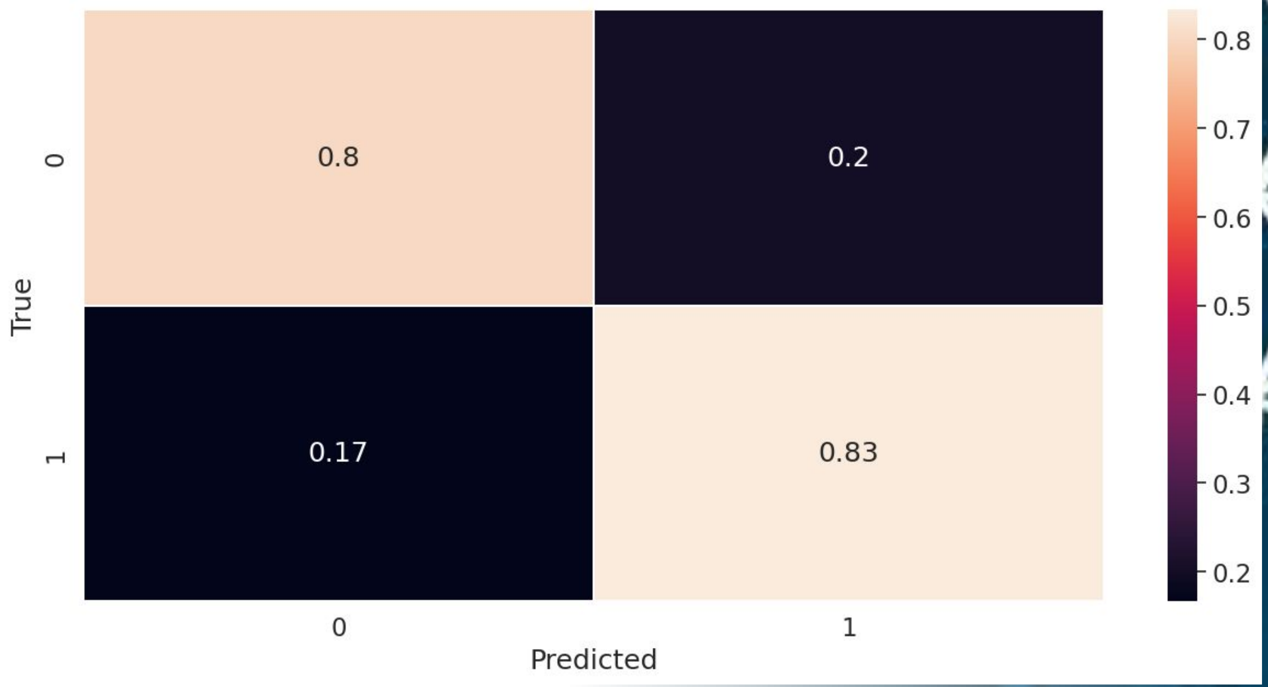
Predictions and evaluation on an external test set after 5 epochs



Results

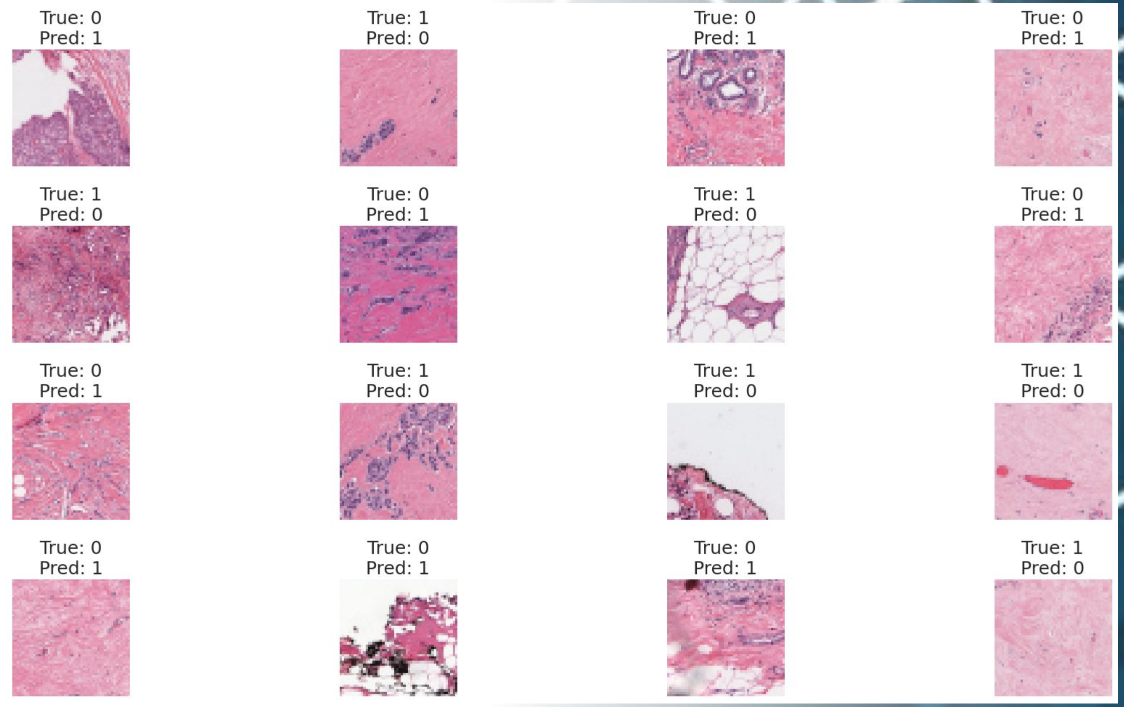
- Performance of EfficientNetV2s on test set

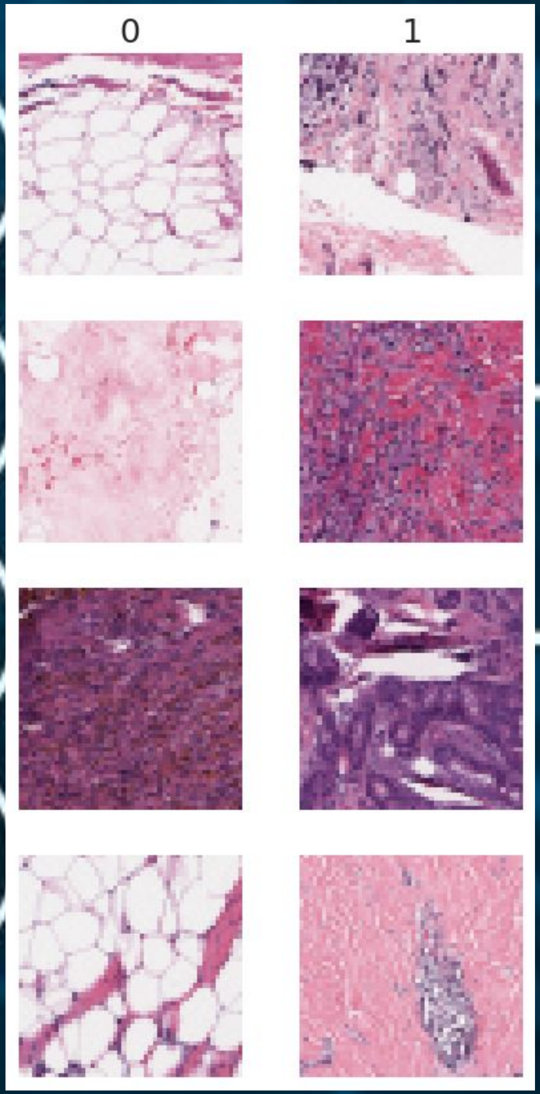
	precision	recall	f1-score	support
0	0.924	0.802	0.858	19874
1	0.625	0.833	0.714	7879
accuracy			0.811	27753
macro avg	0.774	0.818	0.786	27753
weighted avg	0.839	0.811	0.818	27753



Results

- Examples of cancer classification (Attendo immagini del codice runnato)

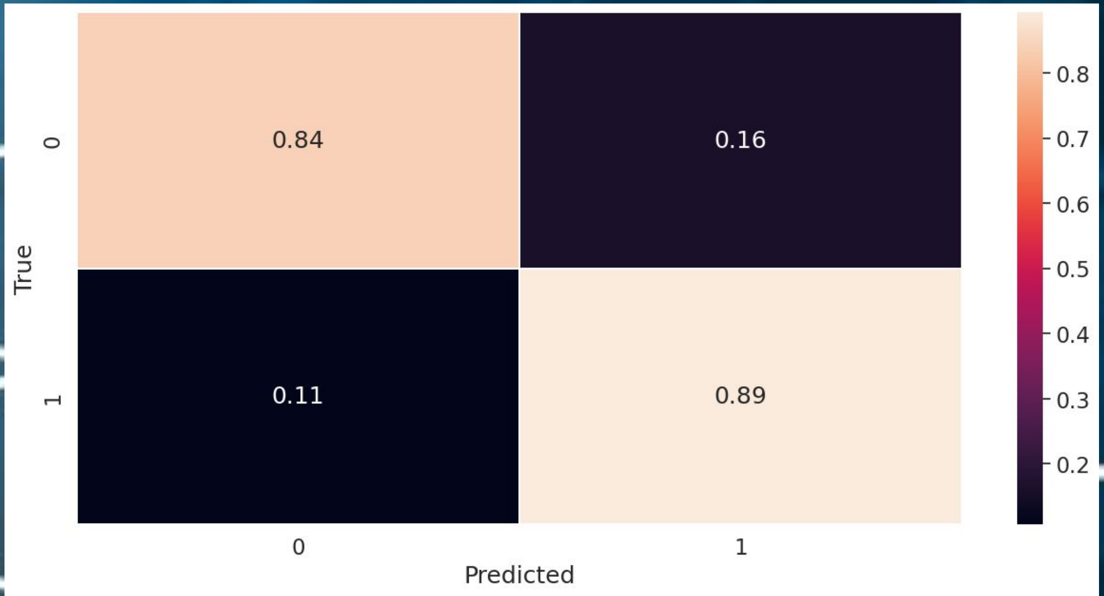


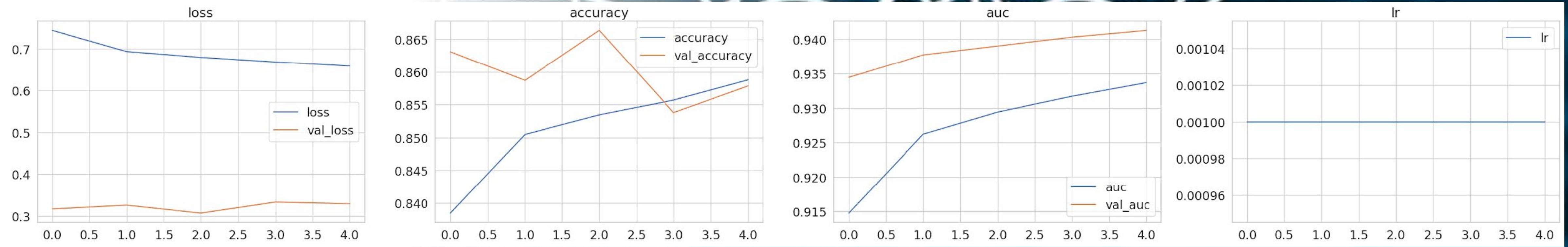


Results

- Performance on test set of ResNet50 trained for # epochs

	precision	recall	f1-score	support
0	0.953	0.841	0.893	4138
1	0.687	0.893	0.777	1622
accuracy			0.855	5760
macro avg	0.820	0.867	0.835	5760
weighted avg	0.878	0.855	0.860	5760





Conclusion (da confermare)

High Accuracy with ResNet50:

ResNet50 demonstrates precision in predicting IDC and non-IDC cases as original project

Improved Robustness:

Results are significantly enhanced with the implementation of a dedicated test set and balanced classes