

# Helicobacter Detection

Vision and Learning,  
Artificial Intelligence, UAB, 2023

Júlia Garcia Torné

Guillem Samper Argelagués

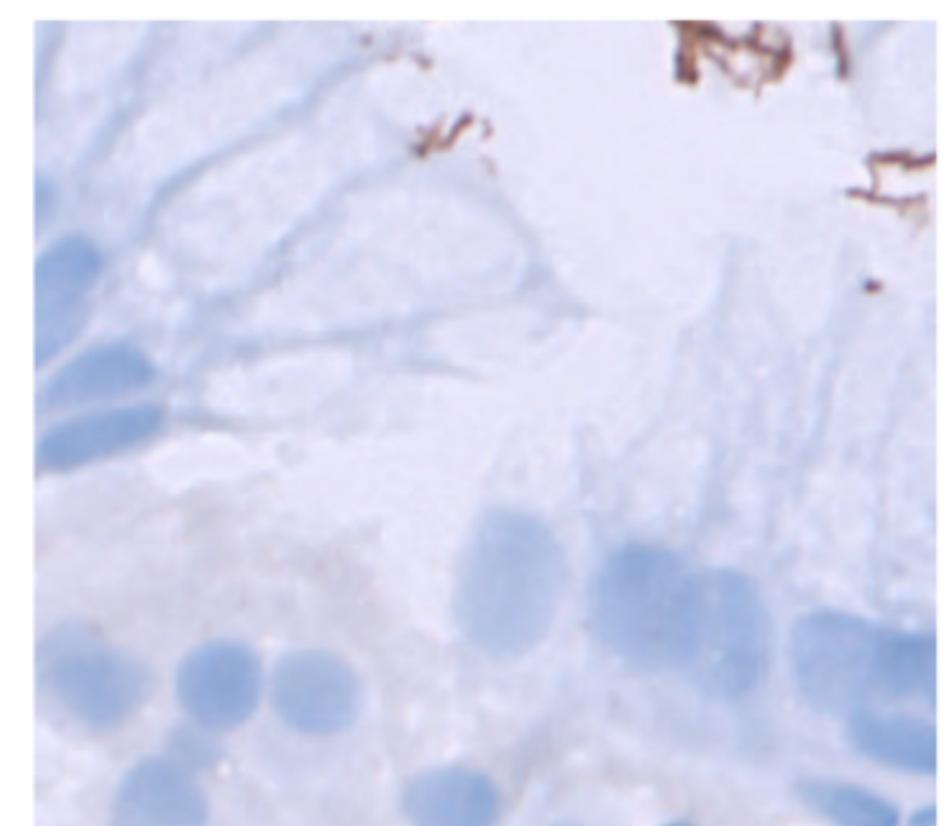
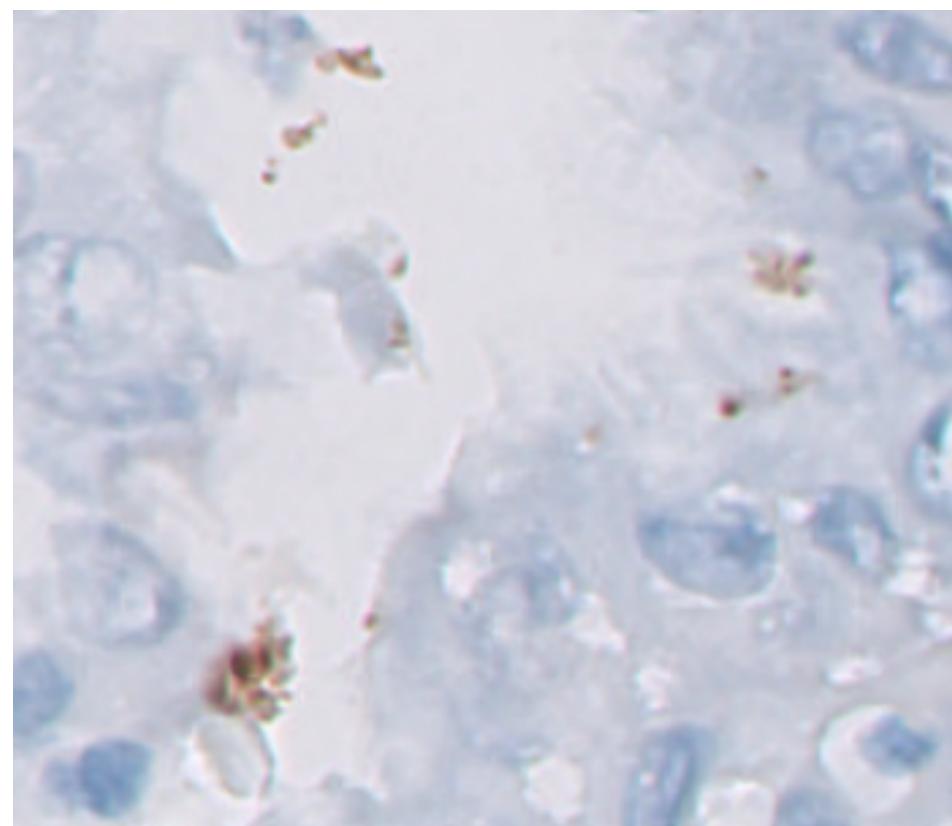
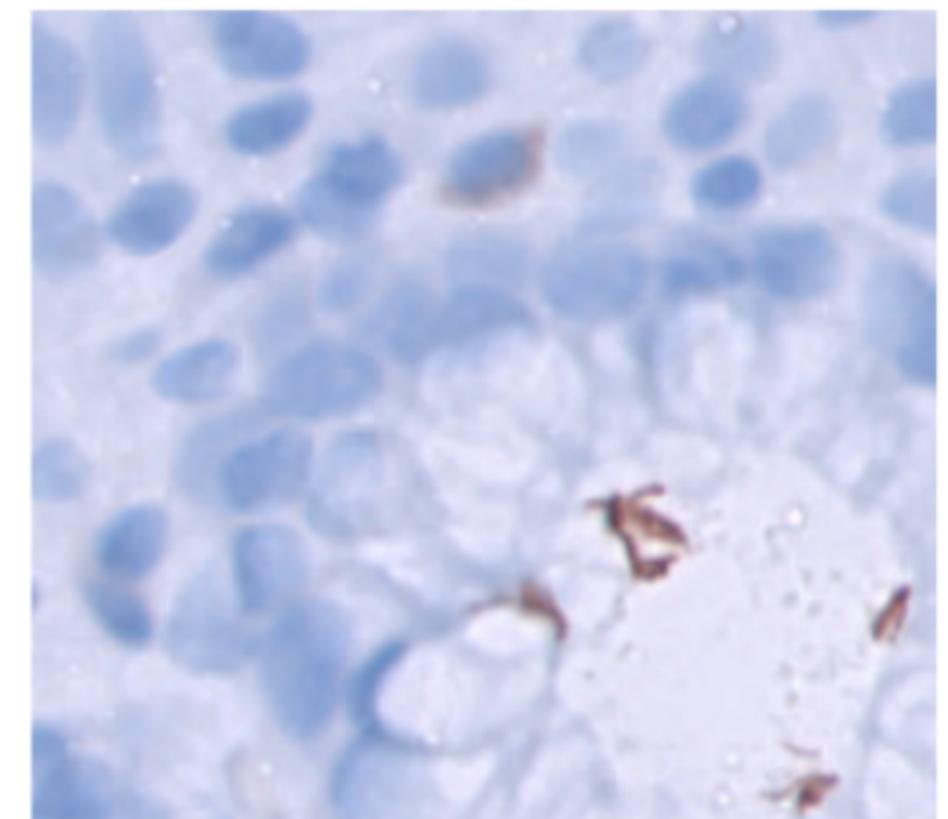
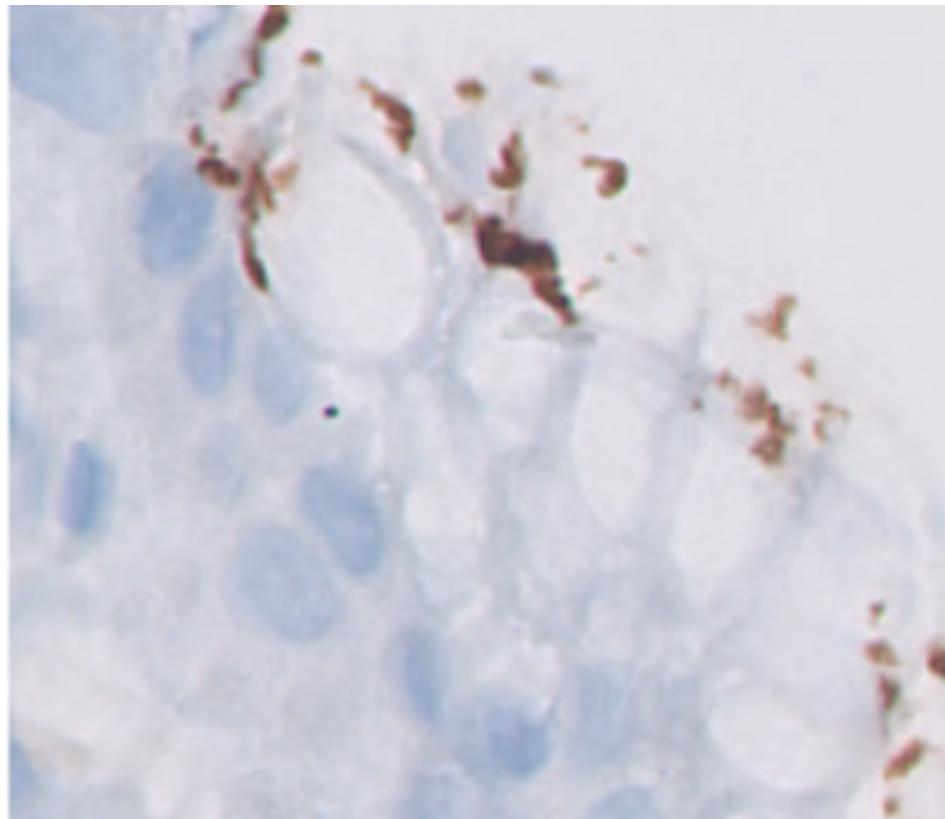
Nerea Qing Muñoz

# CONTENT

- 1 Project Overview
- 2 Dataset Description
- 3 Obtaining the patches
- 4 Model architecture
- 5 Results and performance
- 6 Conclusions

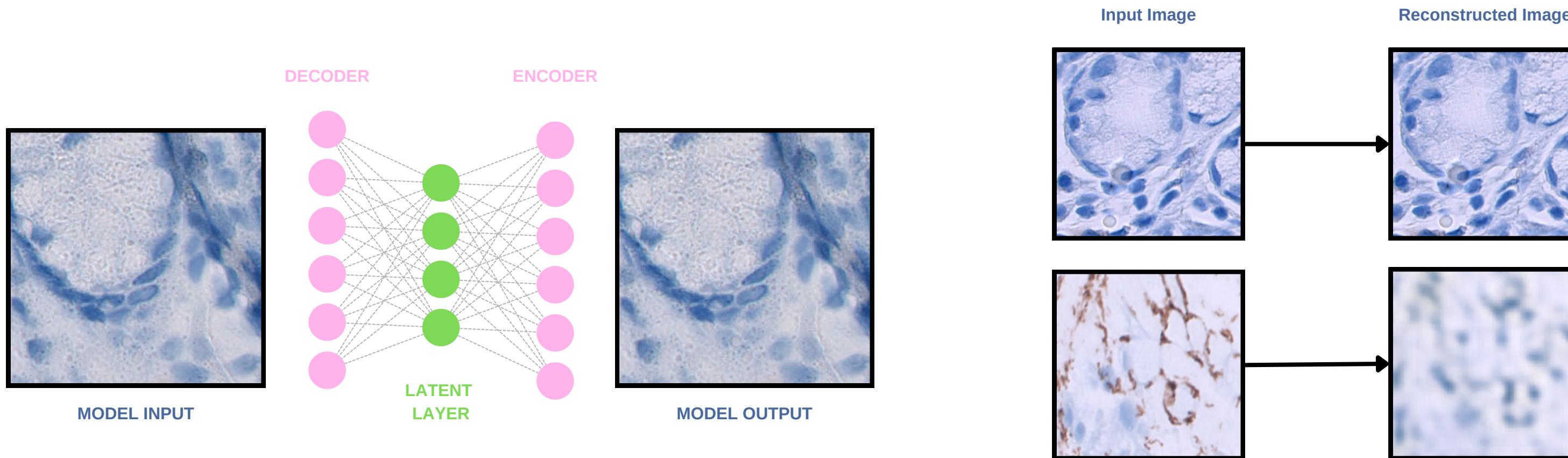
# PROJECT OVERVIEW

Primary Goal: Detecting  
*Helicobacter pylori* in Biopsy  
Image.

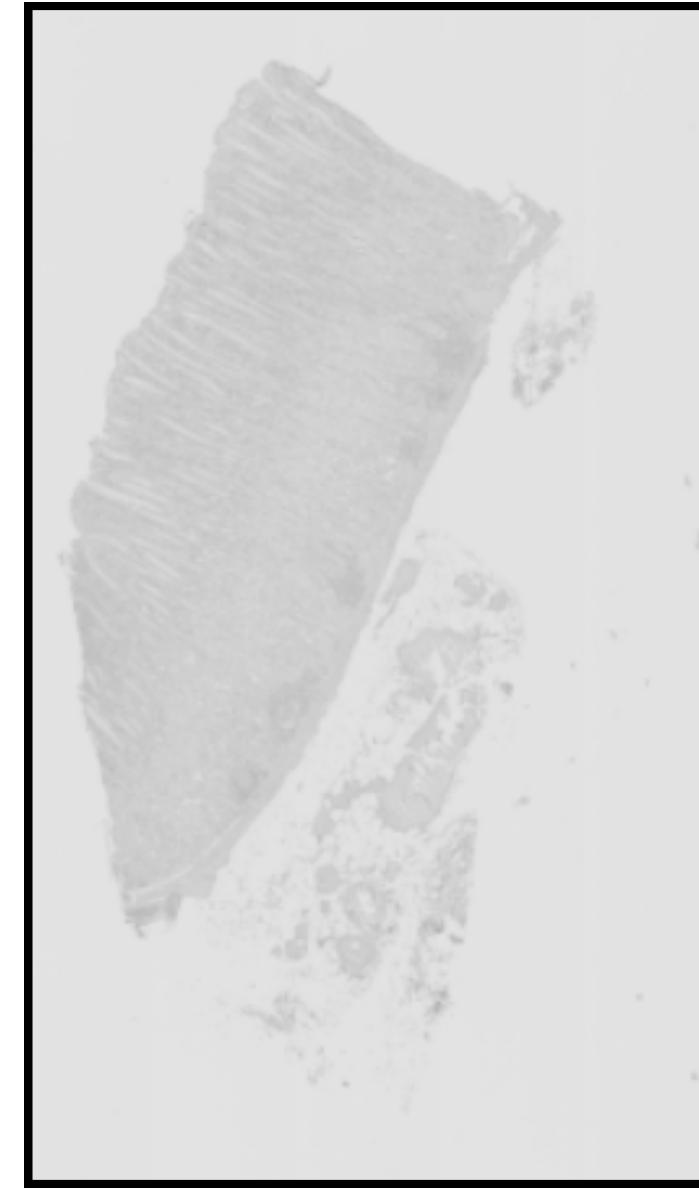


# PROJECT METHODOLOGY

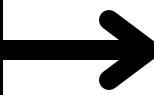
- Create an autoencoder that does a good job at reconstructing samples without helicobacter but struggles when dealing with samples containing the bacteria.
- Employ a pre-defined reconstruction error to differentiate between the two types of samples.



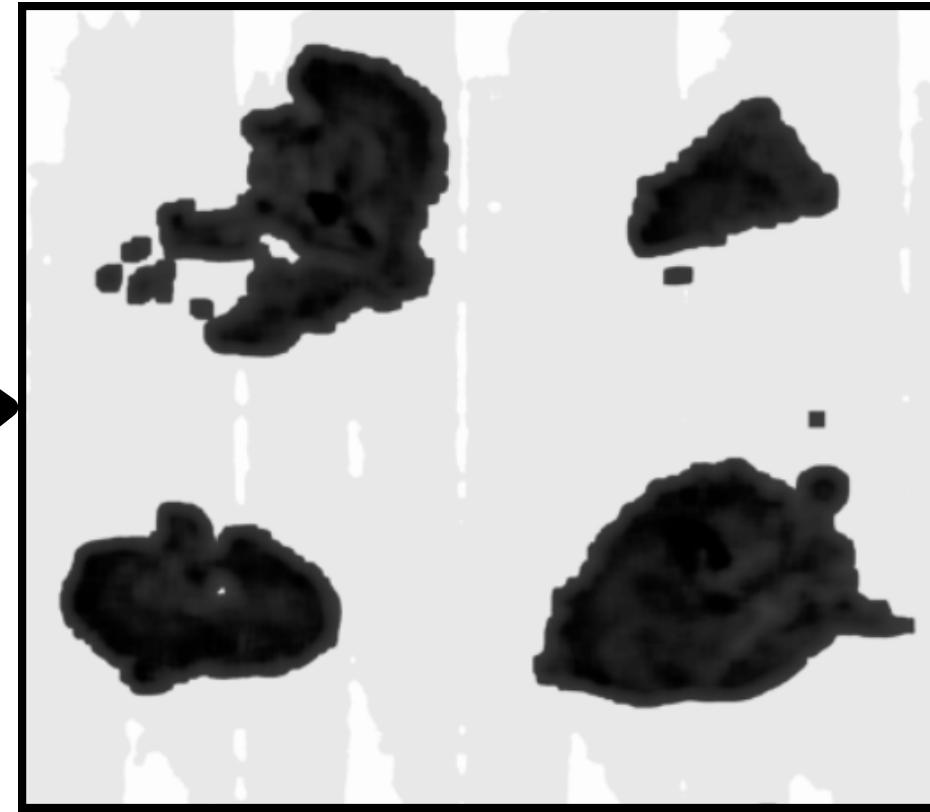
# STEP 1: Obtaining the patches



Grayscale Image



Enhancing Contrast



Gaussian Filter: Reducing Noise

Binary Image

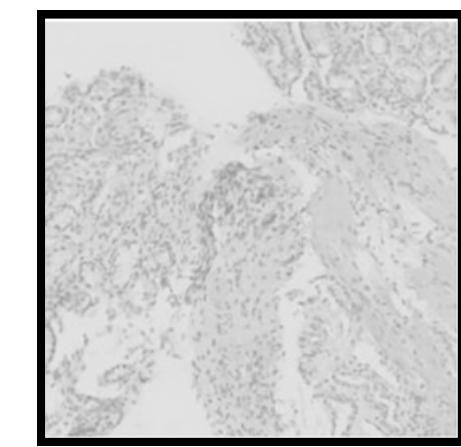
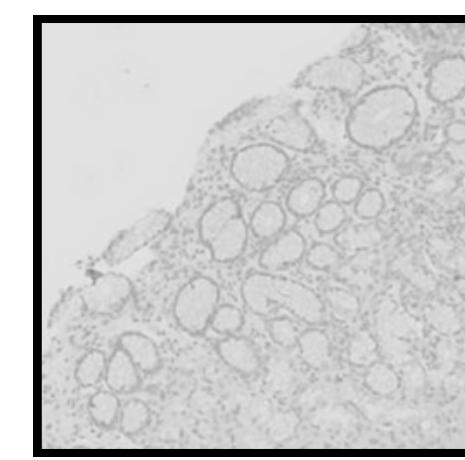
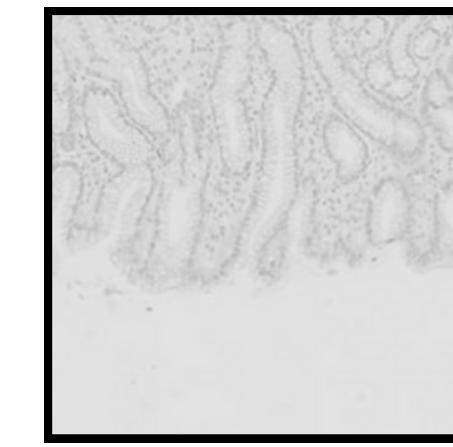
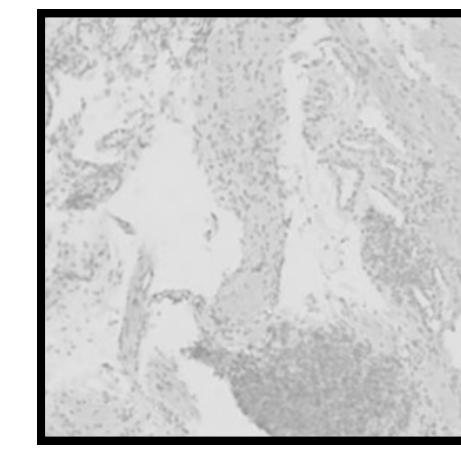
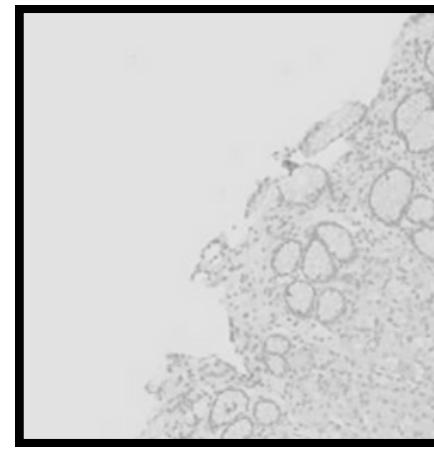
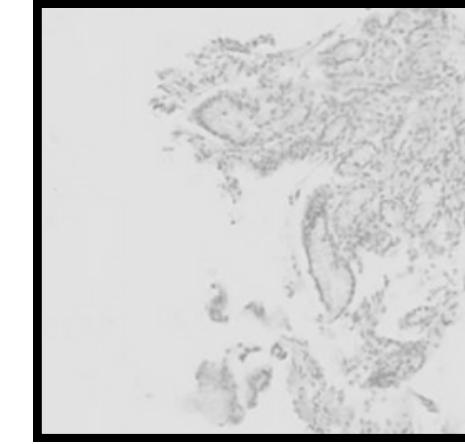
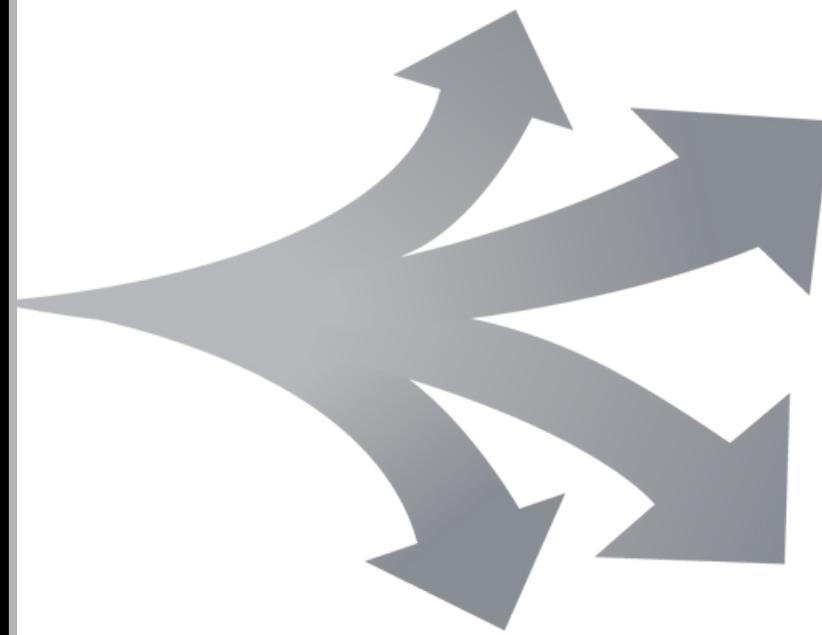


Finding Contours

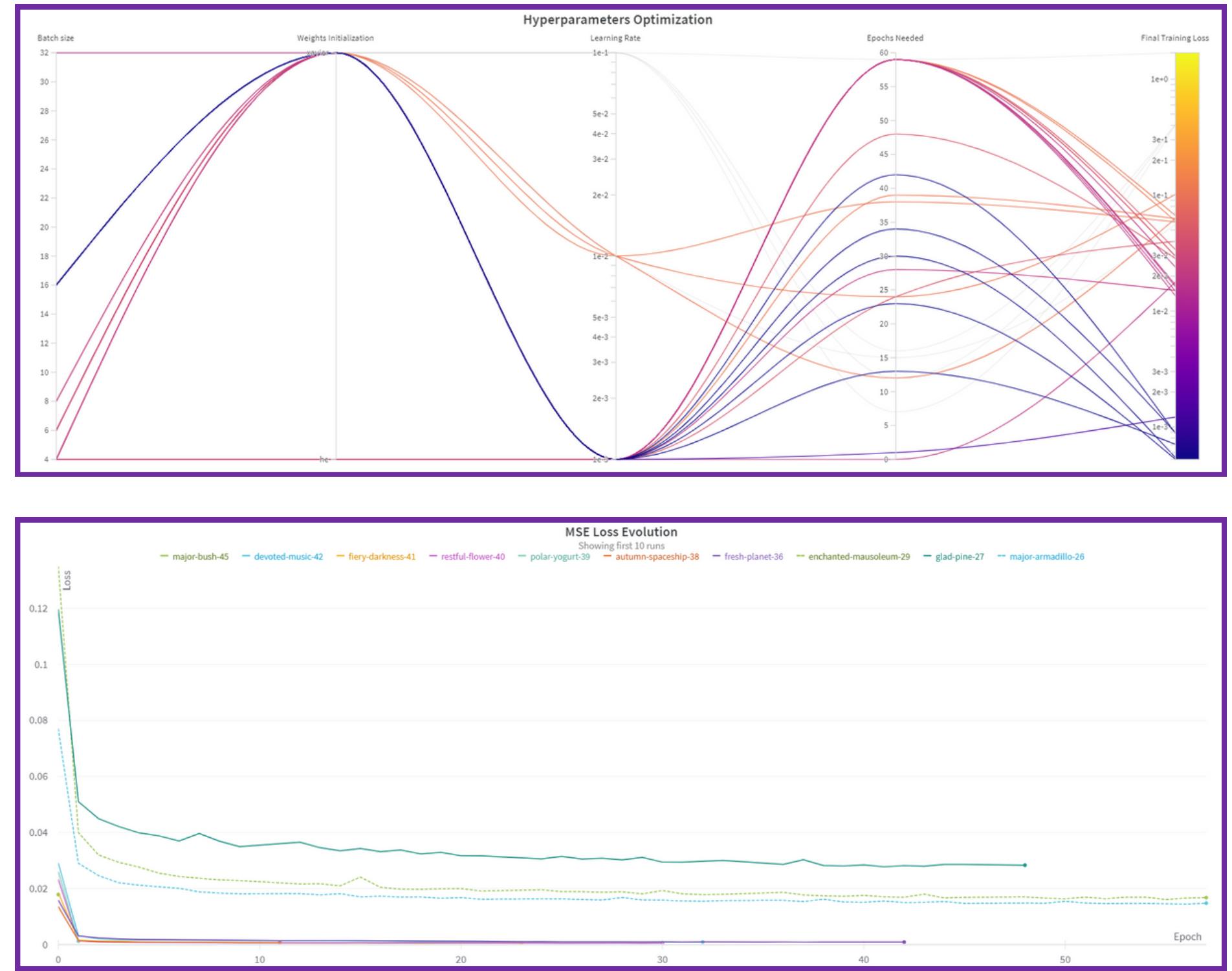


Morphological Operation: Closing





# STEP 2: Finding the best architecture and optimizing parameters



# FINAL MODEL: Training Phase

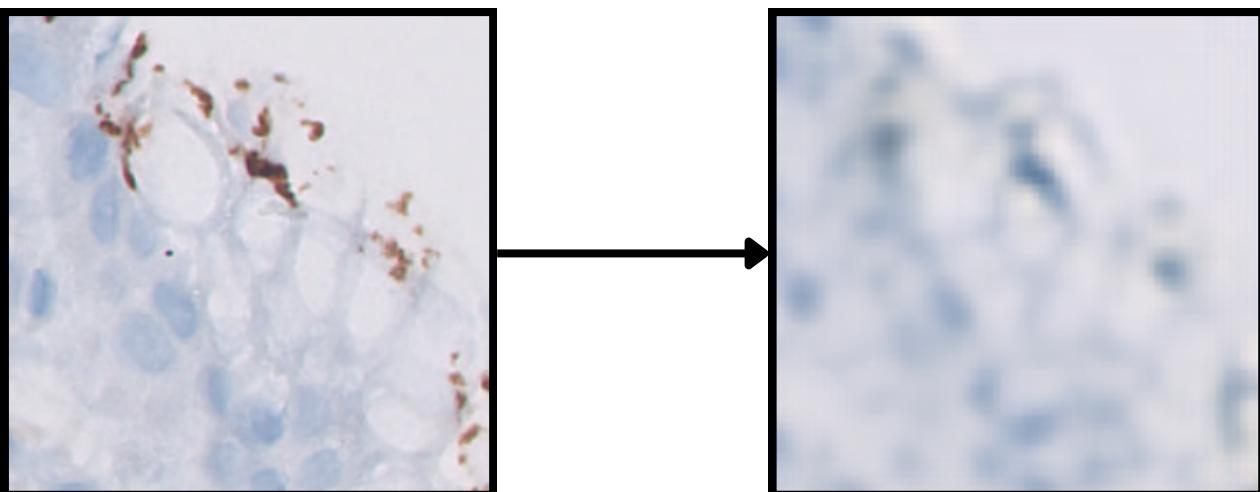
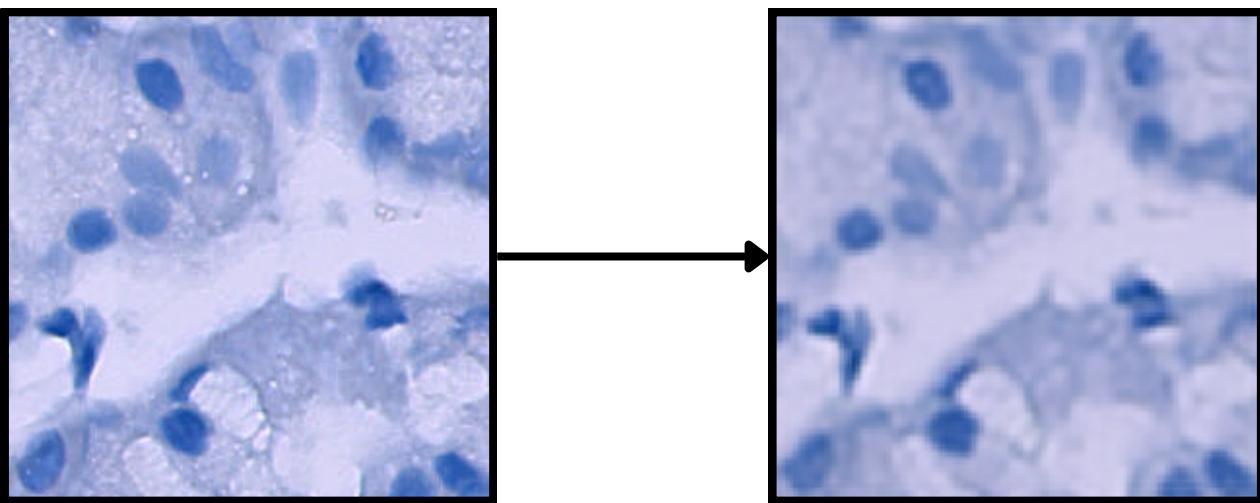


- Batch size: 32
- Optimizer: Adam
- Loss Criterion: Mean Squared Error (MSELoss)
- Learning Rate: 0.001
- Number of epochs: 30
- Architecture:
  - Encoder - [4, 128, 64, 32, 8]
  - Decoder - [8, 32, 64, 128, 4]
- Early Stopping

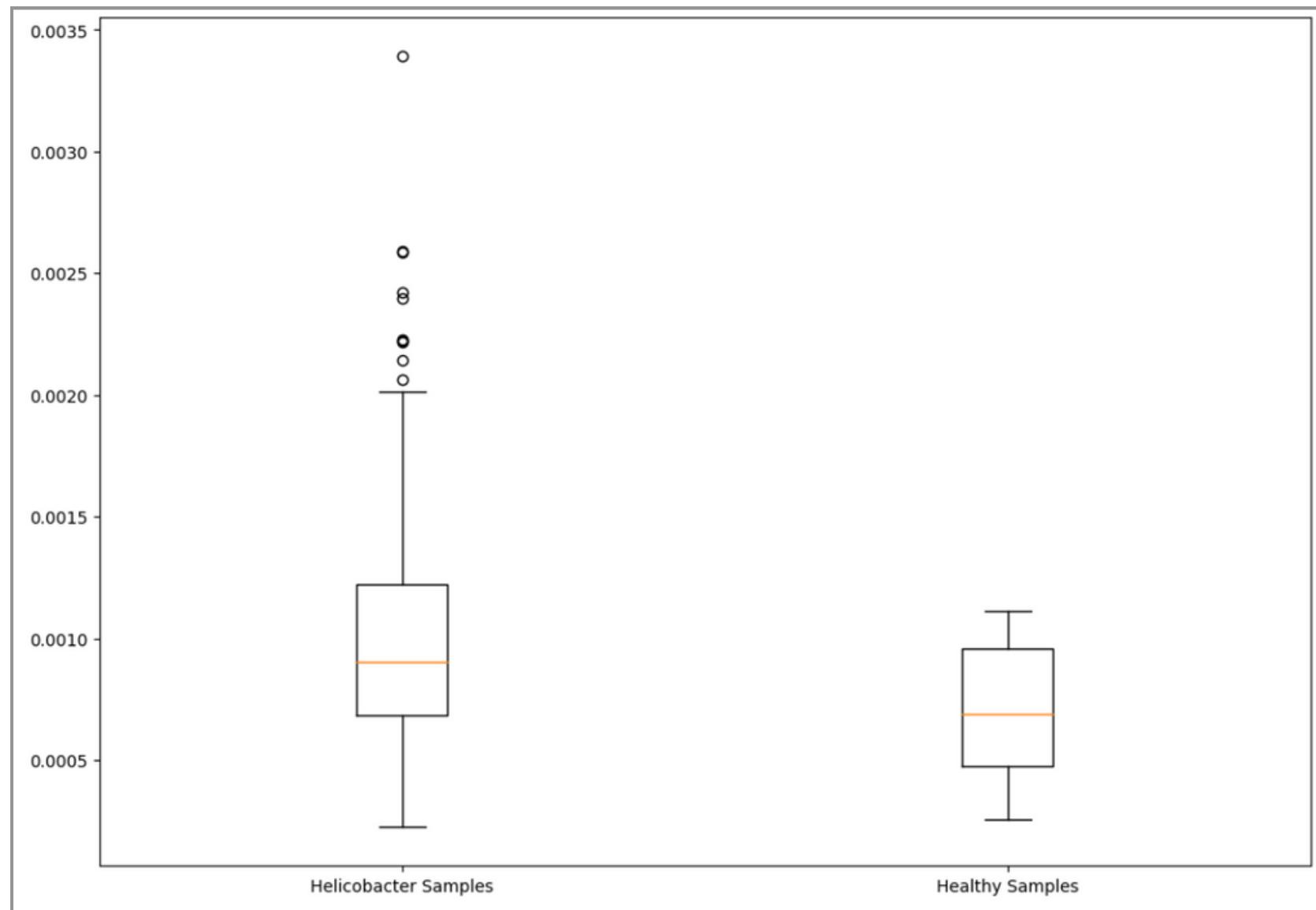
# FINAL MODEL: Analyzing Performance through Visualization

## OBSERVATIONS:

- The model can reproduce images of healthy tissue.
- Since the model was not exposed to Helicobacter during training, it can't reconstruct the **red pixels**.



# EVALUATION: Method 1



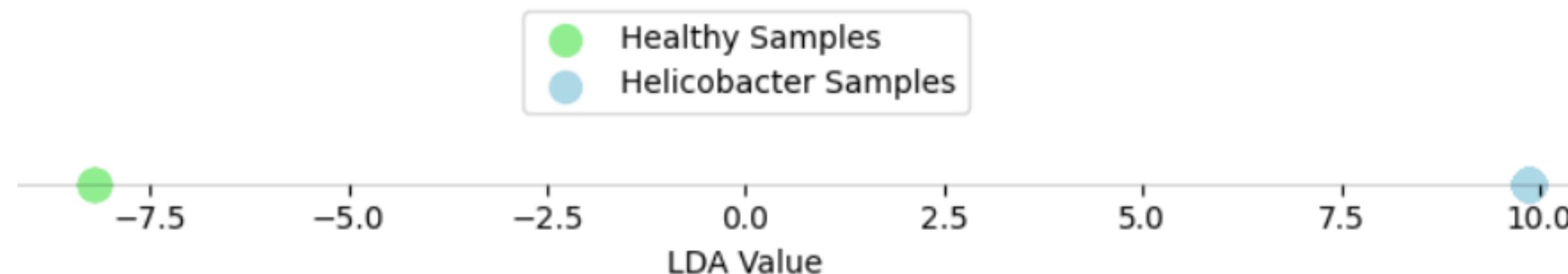
- Reconstruction Error == Mean Squared Error (MSELoss)
- **PROBLEM:** There is no significant difference between the losses for healthy and infected tissues. Not useful for classification.

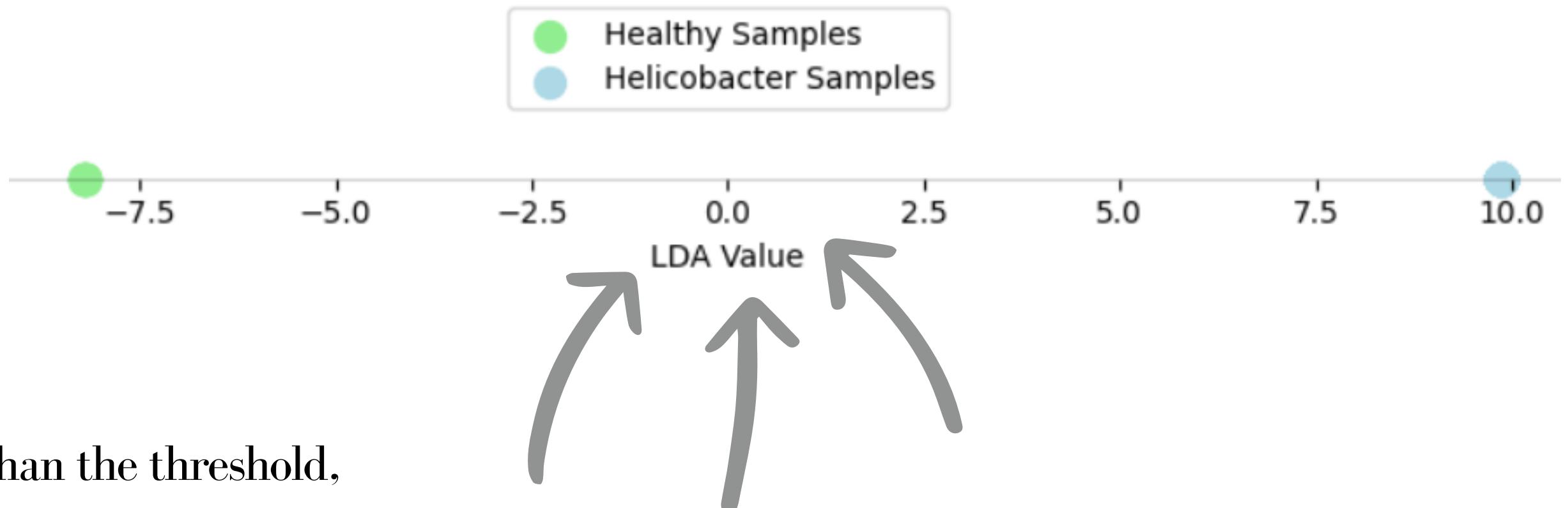
# EVALUATION: Method 2

- **IDEA:** Use the encoded images (vectors residing in the **latent space**) to perform the classification.

## LINEAR DISCRIMINANT ANALYSIS (LDA)

Technique that transforms **high-dimensional vectors** into a **lower-dimensional space**, emphasizing the **separation between classes** for better classification.





THRESHOLD = 0.0

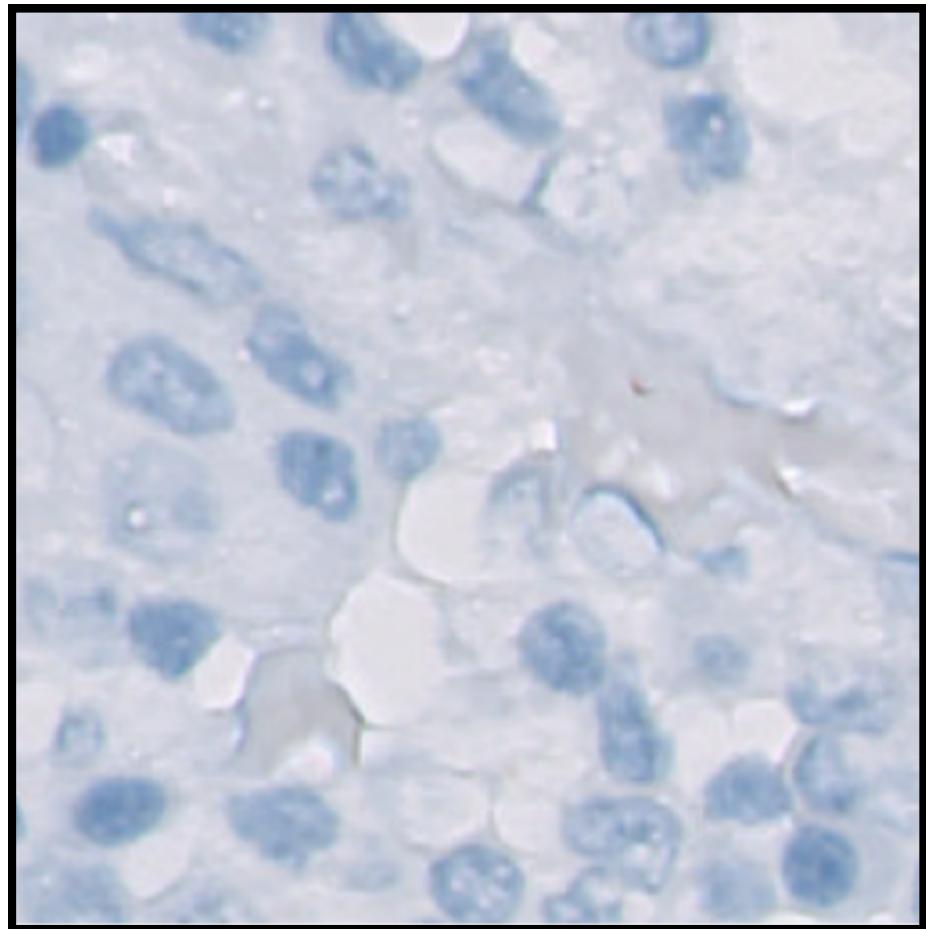
If the value obtained is smaller than the threshold,

the sample will be healthy. Otherwise, it will

contain Helicobacter pylori.

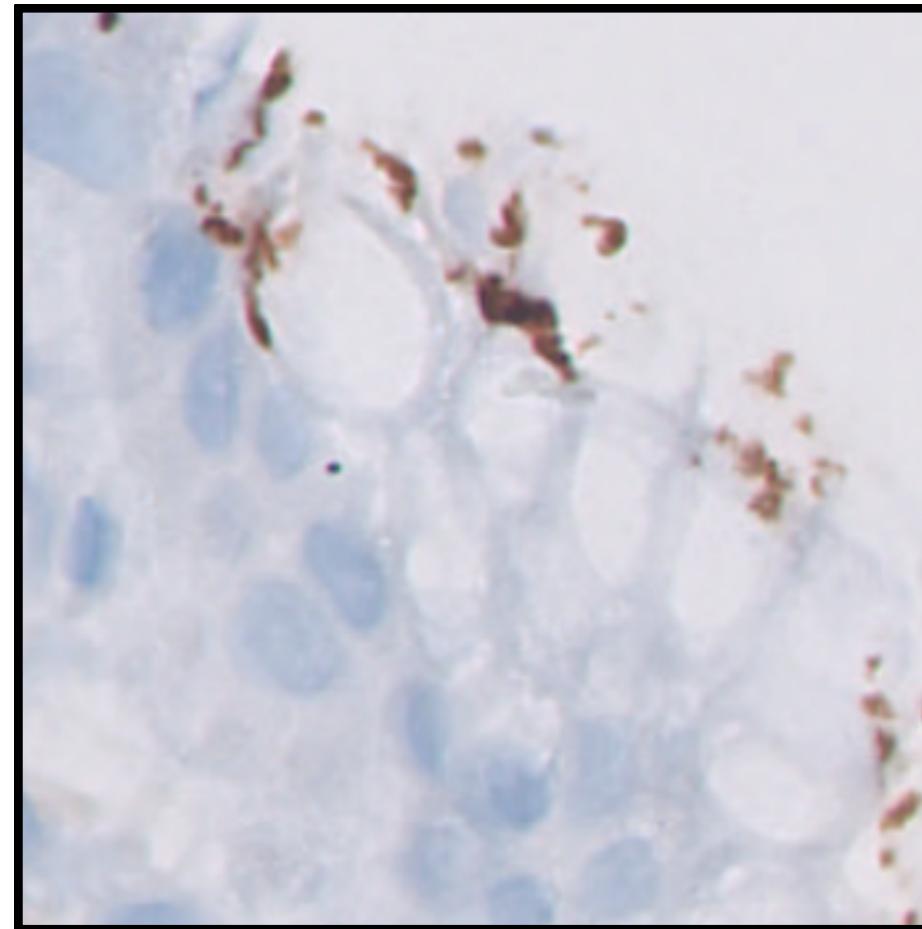
Classified as:

**HEALTHY**



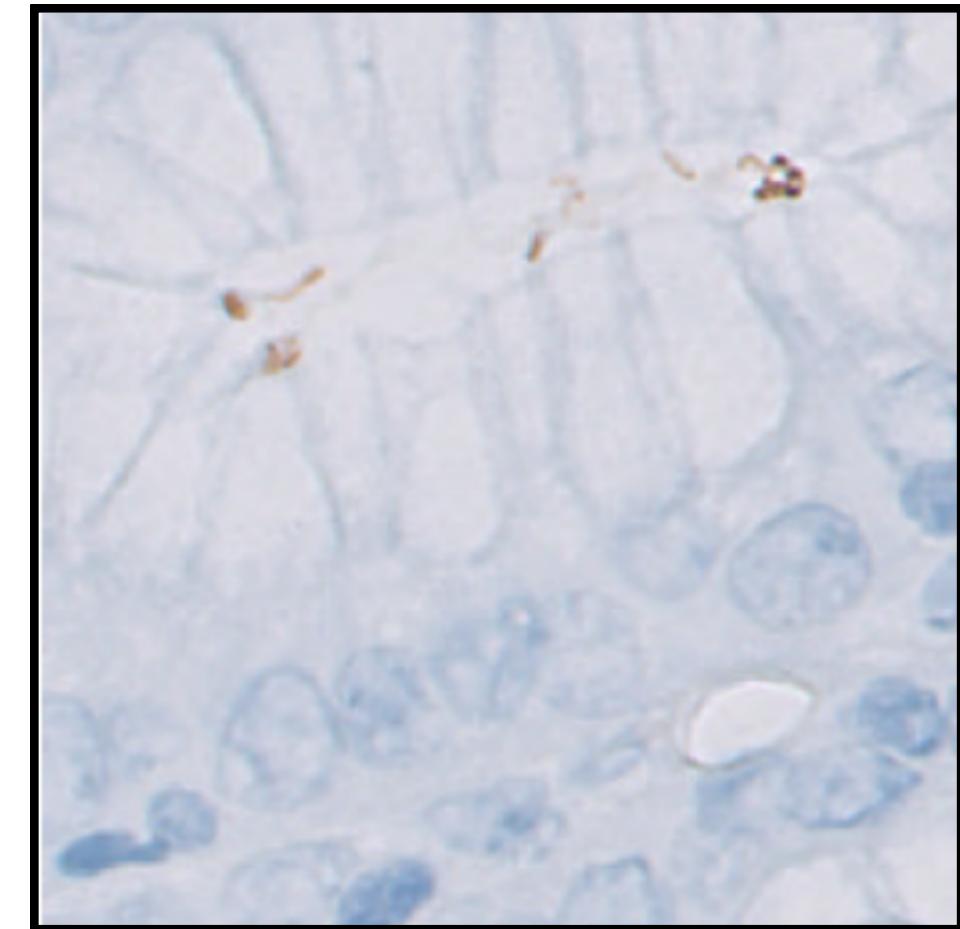
Classified as:

**SICK**



Classified as:

**HEALTHY**

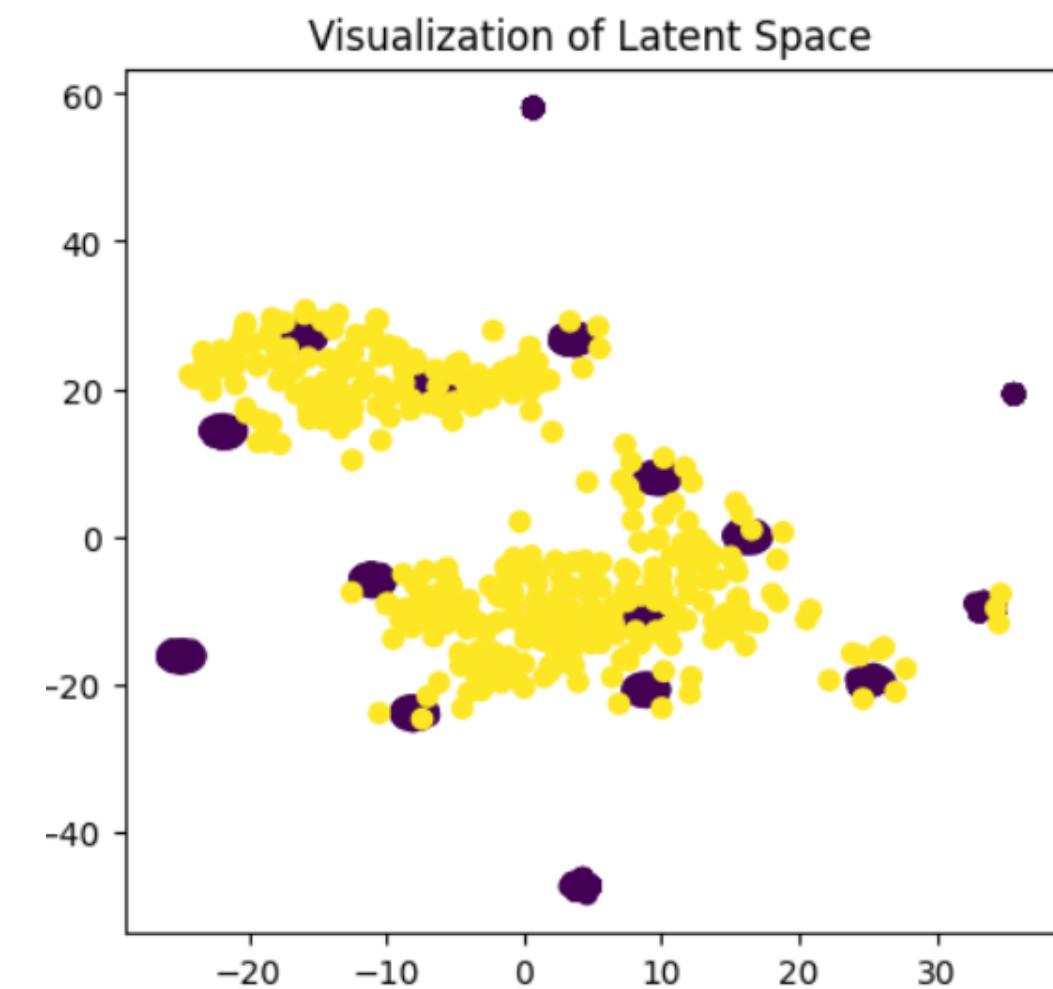


- 9.8491020...

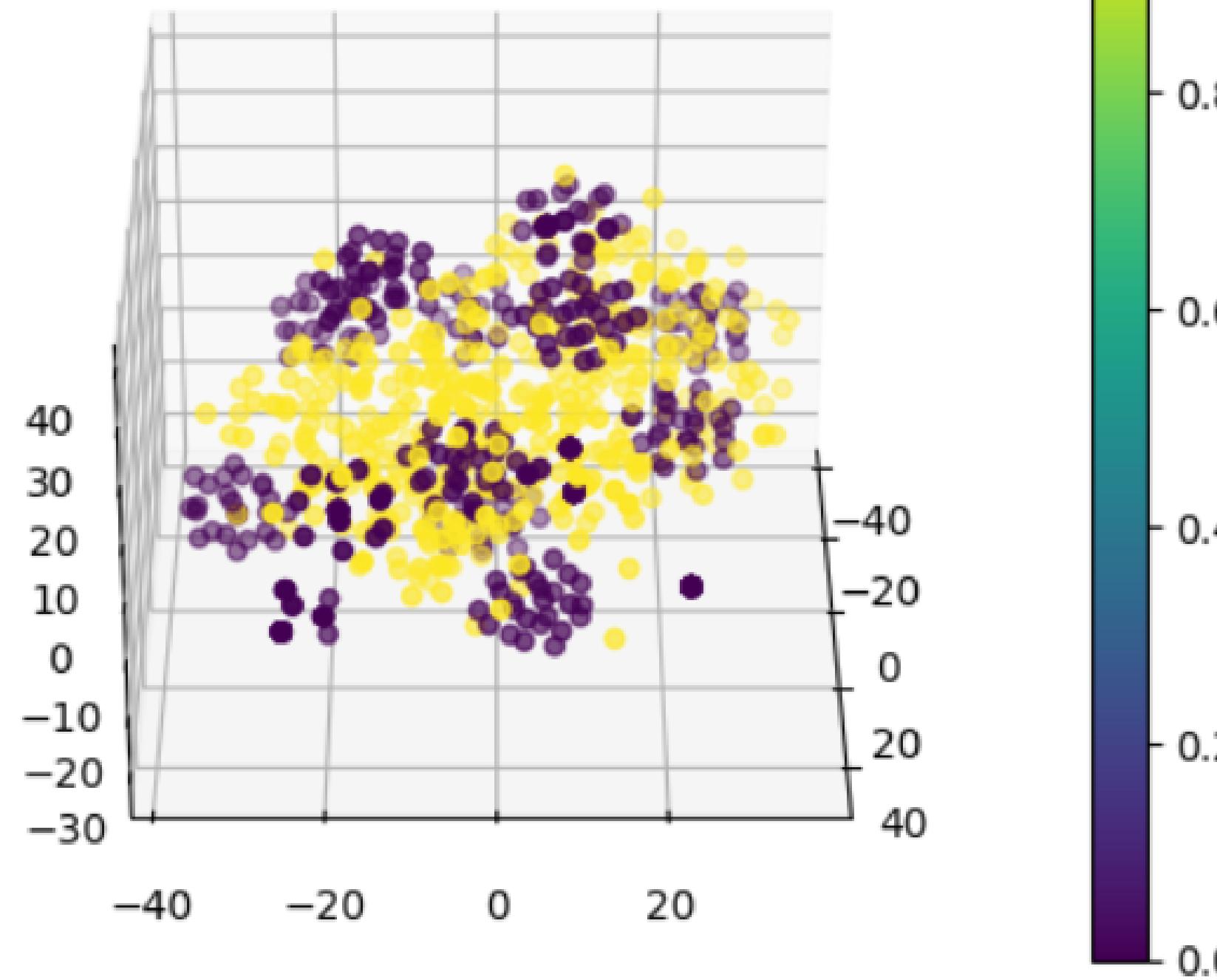
+ 9.8923521...

- 8.2075881...

# T-SNE (T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING)



# 3D Visualization of Latent Space

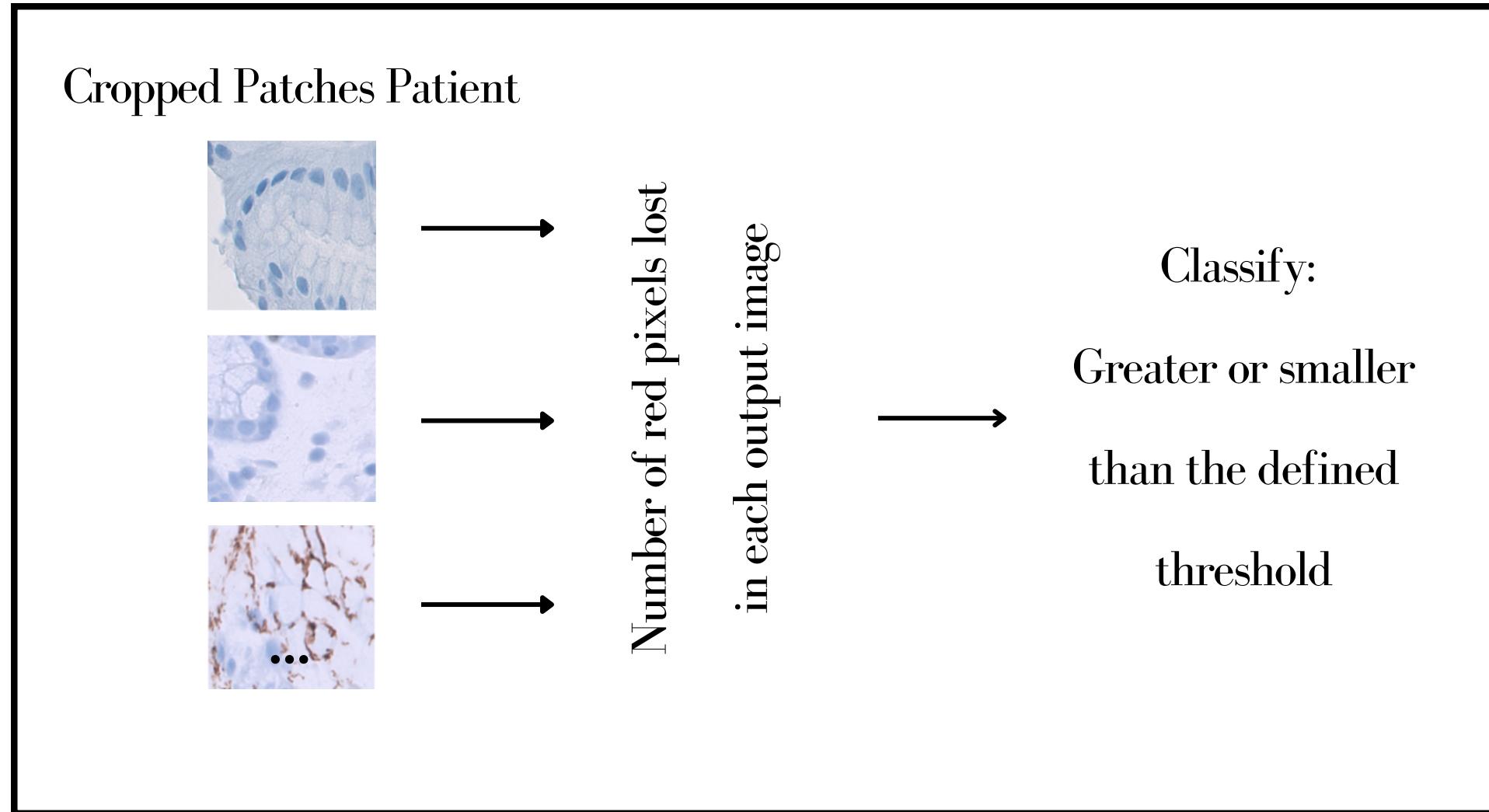


# EVALUATION: Method 3

- IDEA: Given that the model struggles to reconstruct red pixels, use this difference of red pixels between the input and the output images to determine if a window contains the bacteria.

RECONSTRUCTION ERROR = Red pixels in input image - Red pixels in output image

# FINAL EVALUATION: Methodology

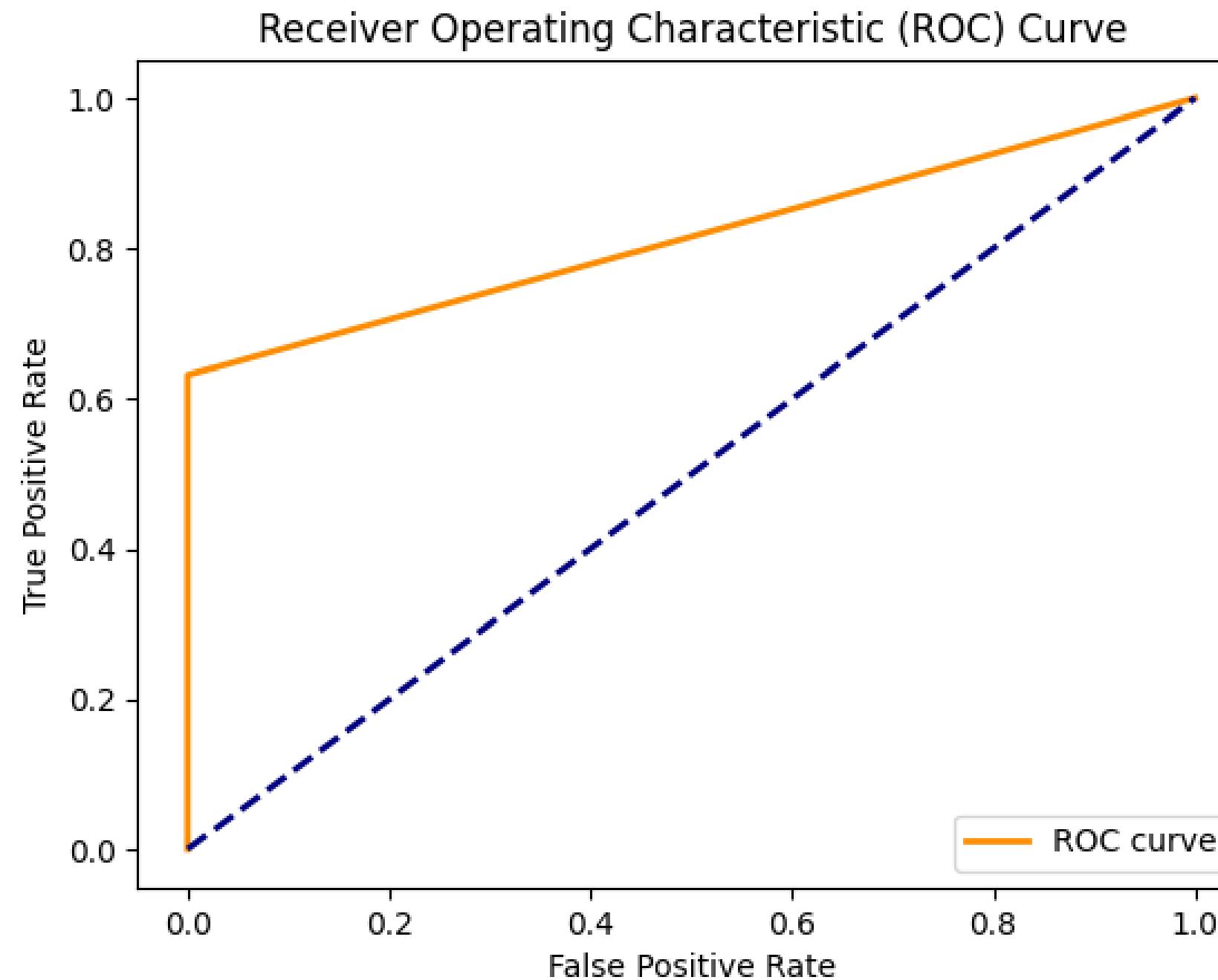


Patient B22-86 has 1255 samples, which 322 of them contain helicobacters.

Percentage of helicobacters: 25.65%

Result: Positive

# FINAL EVALUATION: Definig the threshold



Accuracy: 0.74

Precision: 1.0

Recall: 0.63

F1 score: 0.77

# CONCLUSIONS

- Good performance of the autoencoder
- Different methodologies: difference of red pixels is the best one
- Validation not fully completed