Heliobacter pylori Diagnosis

Marek Tutka, Nico Enghardt, Bogdan Mateescu

CONTENTS

GOALS

IMPROVE ACCURACY

Enhance the precision and reliability of bacteria detection and minimize false positives and negatives

EFFICIENCY

Faster diagnoses which enable us to work on higher volumes of images

ADAPTABILITY

Ensure that the model is robust and performs well across new patients never seen during training

KEY CHALLENGES

IMAGE INCONSISTENCIES

Histological images may have different staining intensities or background noise, overlapping colors and color differences

VARIOUS FORMS

H. plyori can appear in different shapes and sizes, some are hardly visible, causing false positives or false negatives

LOW NUMBER OF PATIENTS

A small patient set is sensitive to biases in classes

PatchClassifier MODEl (1)

To be able to find bacteria in patches, we train a custom model, validate it and then evaluate it

PatchClassifier MODEL (2) - LAYERS

CONVOLUTIONAL LAYERS

For extracting features from the input images

MAX-POOLING LAYERS

These downsample the feature maps

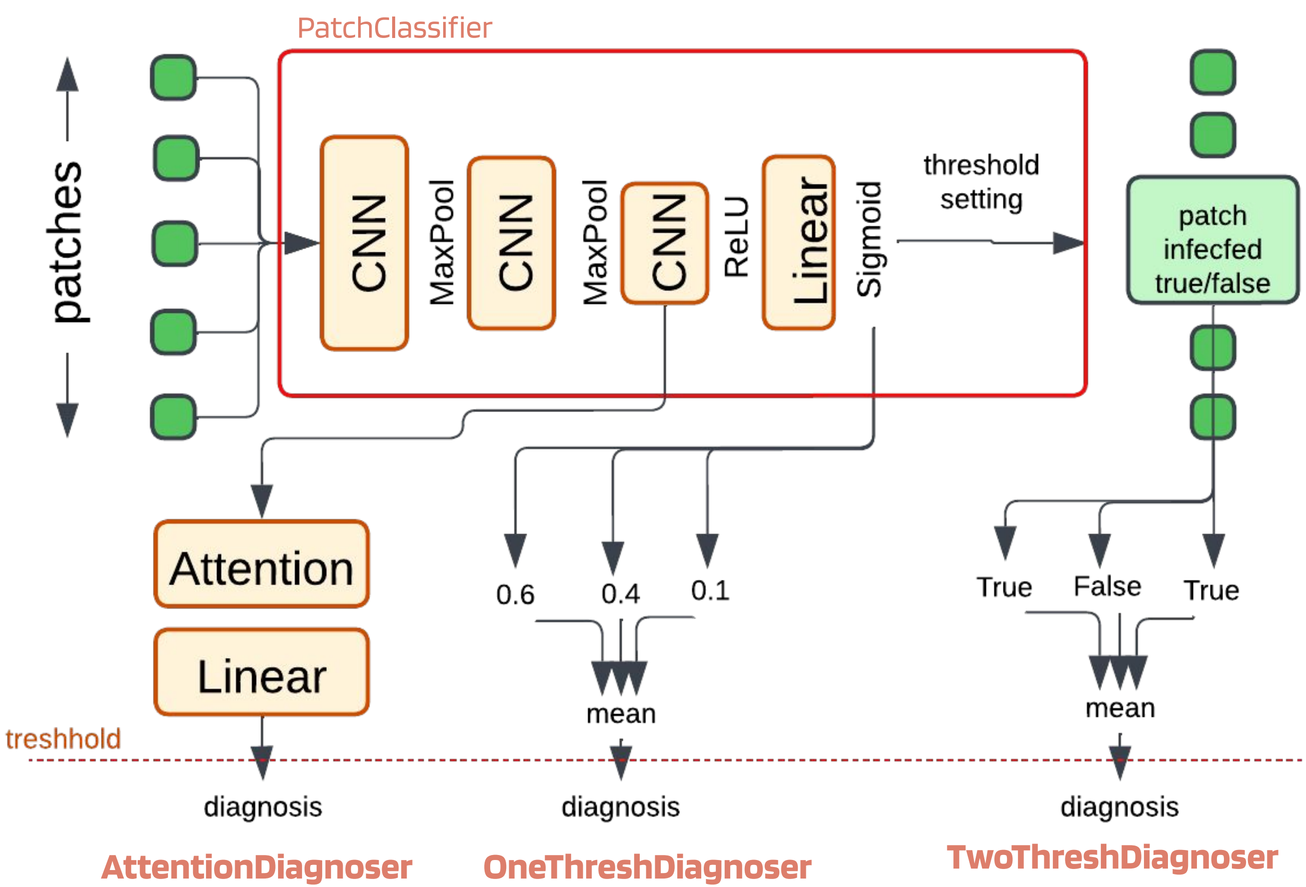
LINEAR LAYER

For providing a probability score ranging from 0 to 1

PATIENT DIAGNOSIS (1)

Adaptive thresholding:

- One Threshold Diagnoser
- Two Threshold Diagnoser
- Attention + Thresholding


PatchClassifier
patches
CNN
MaxPool
CNN
MaxPool
CNN
ReLU
Linear
Sigmoid
threshold setting
patch infecfed true/false
Attention
Linear
0.6
0.4
0.1
True
False
True
mean
mean
treshhold
diagnosis
diagnosis
diagnosis
AttentionDiagnoser
OneThreshDiagnoser
TwoThreshDiagnoser

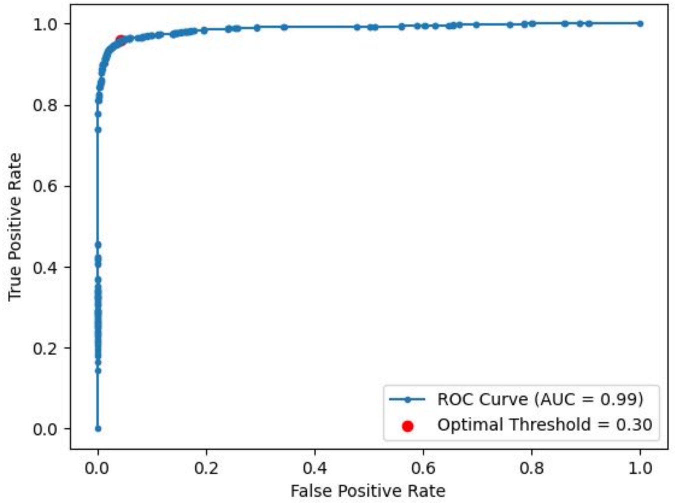
PATIENT DIAGNOSIS (2) - ADAPTIVE THRESHOLDING

One threshold:

- Patches have value from 0.00 to 1.00 and patient's score is a mean value of all patches

Two thresholds:

- Binarize patch scores (either 0 or 1)
- Patient's score is a percentage of sick patches



ROC - Patch classifier

OPTIMIZATION TECHNIQUES (1) – EARLY STOPPING

A way to set a higher number of epochs without overfitting the model

DATASET

“Annotated” dataset – set of patches 3051 patches annotated by the expert. On this data-set we are doing the training

“Cropped” dataset - set of patches for 157 patients, however in this dataset we have expert's diagnosis only for the patient, not for the individual patches. This dataset is being used for evaluation

“Holdout” dataset - Independent set of 120 patients, used for verification of reproducibility.

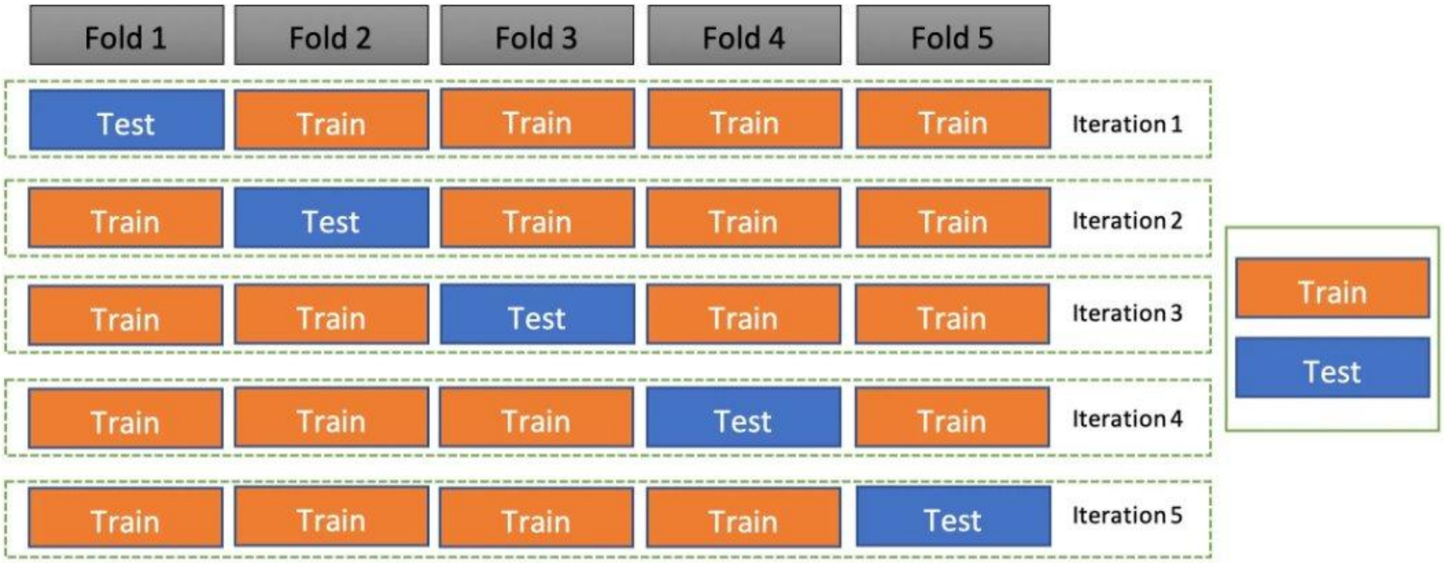
DATASETS

- Split Patients → Train/Test: 5-Fold

- Annotated Patches → train and test **PatchClassifier**
- Cropped Patches → run PatchClassifier and train **Diagnoser**

- HoldOut Patients → evaluate Diagnosers

OPTIMIZATION TECHNIQUES (2) - K-FOLD VALIDATION



EVALUATION – PatchClassifier

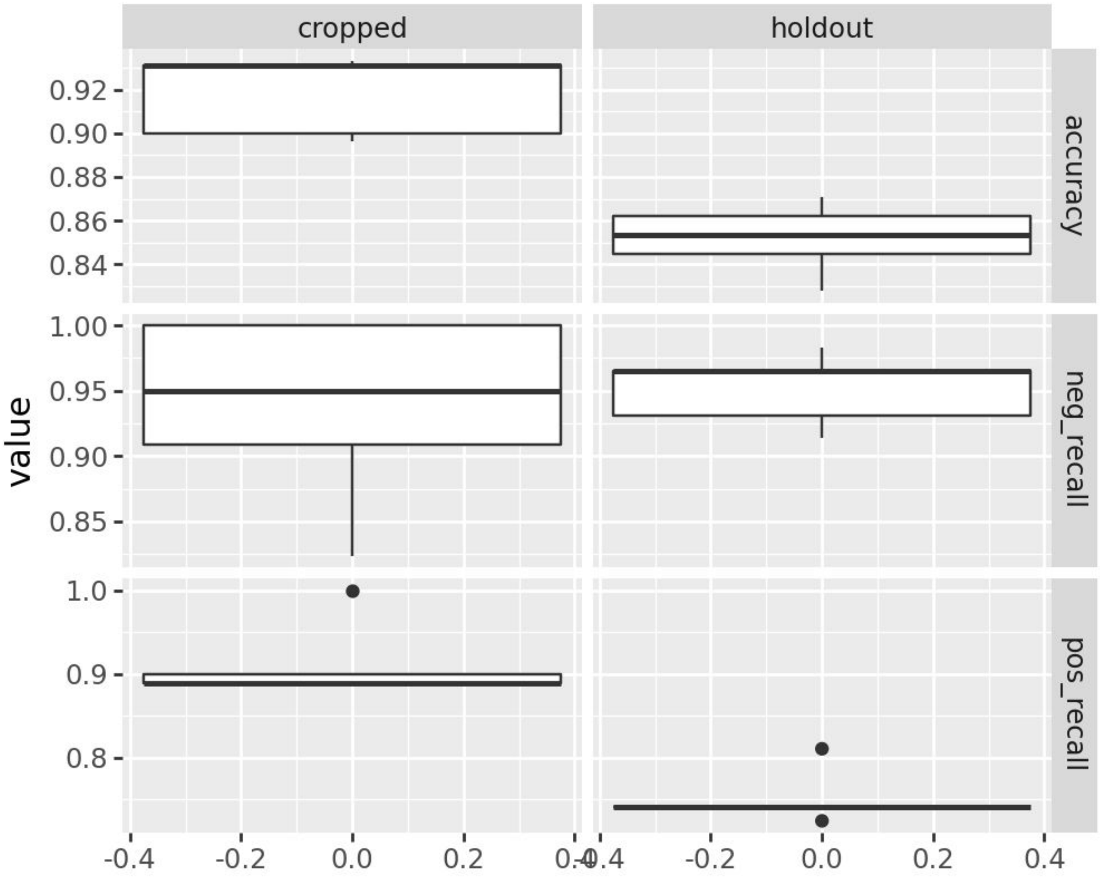
AVERAGED OVER 5 FOLDS	PatchClassifier
Accuracy	94 % ± 4%
Precision	95% ± 2 %
Positive Recall	93% ± 7 %
Negative Recall	95% ± 3%
F1	94% ± 3%

Averaged over 5 folds	Neg Pred	False Pred
Neg Groundtruth	53 % ± 12 %	3,5 % ± 3,5 %
Positive Groundtruth	2,4 % ± 1,1 %	41 % ± 10%

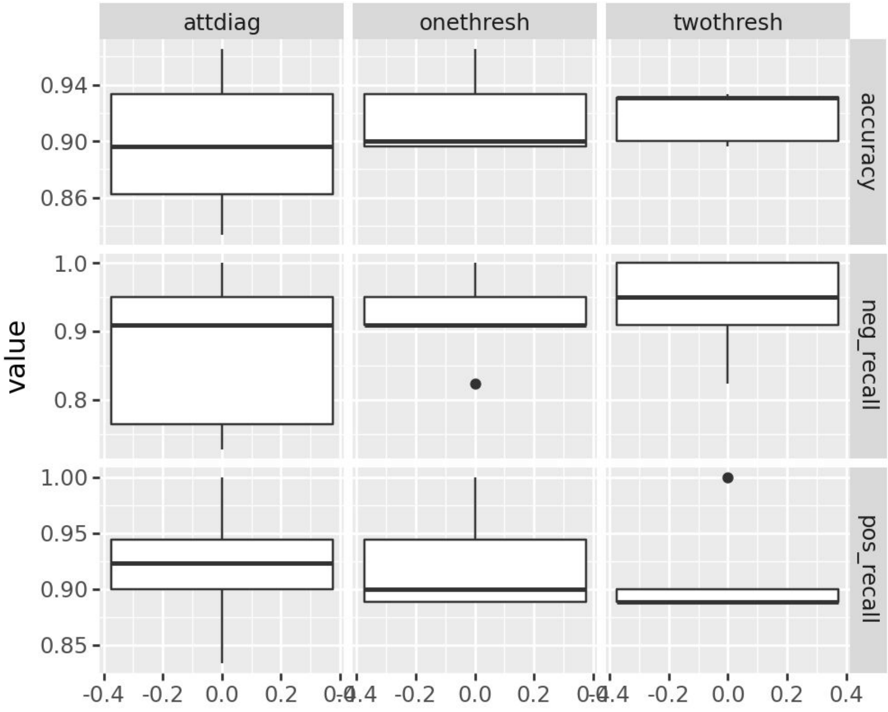
DIAGNOSER SCORES

TEST / HOLDOUT	AttentionDiagnoser	OneThresh Diagnoser	TwoThresh Diagnoser
Accuracy	90% ± 5% / 84% ± 2%	92% ± 3% / 85% ± 2%	92% ± 2% / 85% ± 1%
Precision	89% ± 8% / 95% ± 3%	92% ± 6% / 93% ± 3%	93% ± 7% / 94% ± 3%
Positive Recall	92% ± 5% / 73% ± 4%	92% ± 4% / 76% ± 3%	91% ± 4% / 75% ± 3%
Negative Recall	87% ± 11% / 96% ± 3%	92% ± 6% / 94% ± 3%	94% ± 7% / 95% ± 3%
F1	90% ± 5% / 82% ± 2%	92% ± 3% / 83% ± 2%	92% ± 2% / 84% ± 2%

CROPPED SET VS HOLDOUT SET



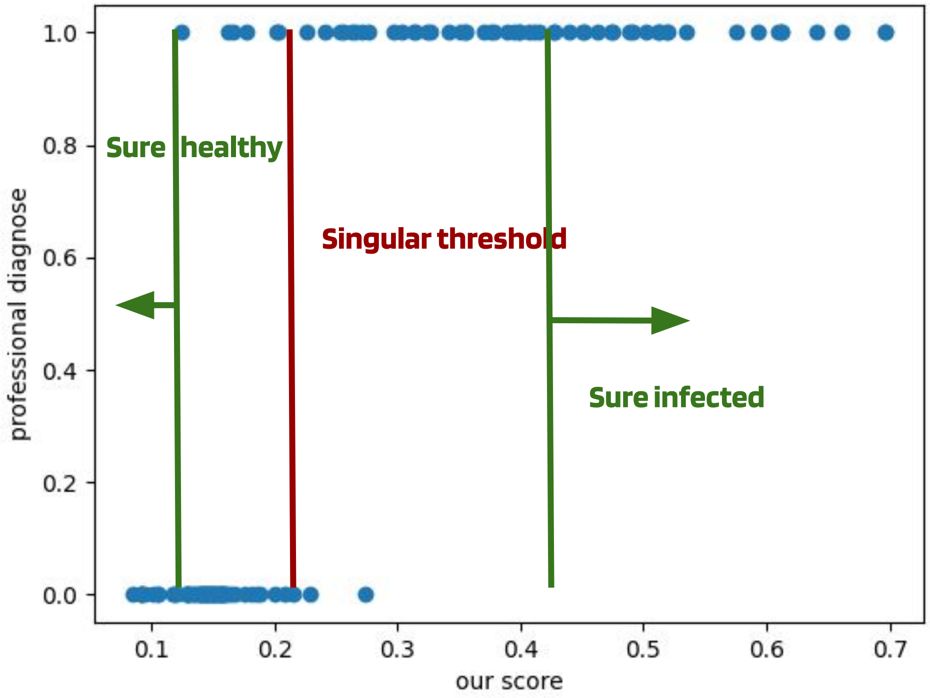
AttentionD. VS OneThresh. VS TwoThresh.



Further Ideas:

Healthy/Unsure/Sick

Cooperation between System and Doctor



THANK YOU

Any questions?