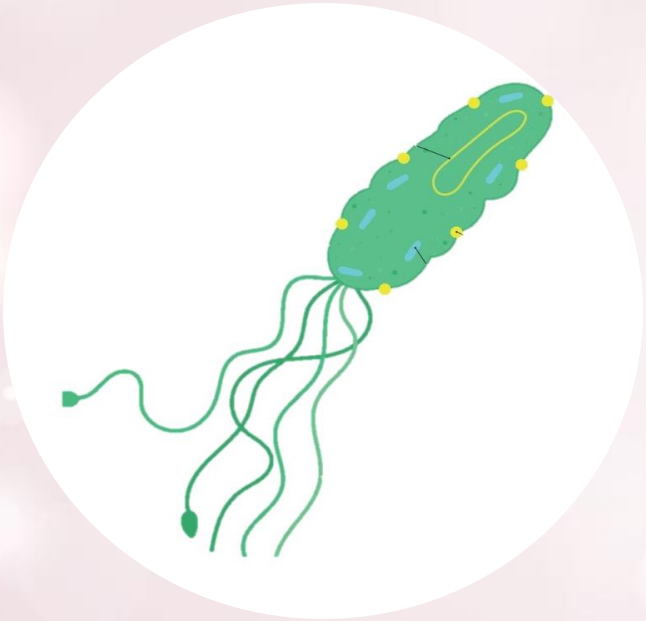
Machine-Learning-Guided Diagnosis for Heliobacter Pylori

Katharina Alexa Lang & Nico Enghardt, 7. Jan 2024

Outline

- Motivation
- Convolutions in Image Processing
- Convolutional Neural Networks
- Results



Helicobacter Pylori

- Bacteria colonizes stomach
- 40% global incidence
- Often asymptomatic
- 10-20% develop ulcerus
- 0.5-2% develop stomach cancer

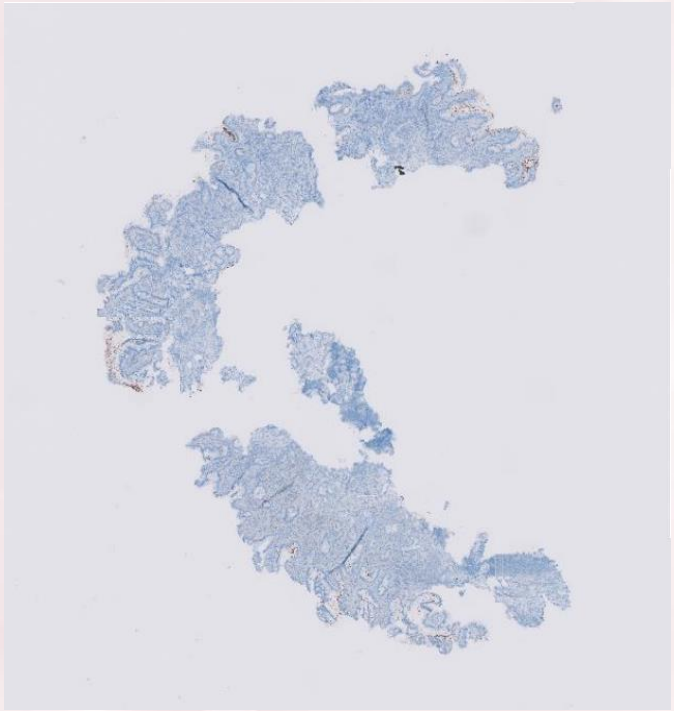
Diagnosis of H.pylori

Use case:
symptoms of stomach infection
or high genetic risk of stomach cancer

Methods:
Stomach biopsy or antibody detection in blood

Treatment:
Patients receive antibiotics

Biopsy analysis is time consuming

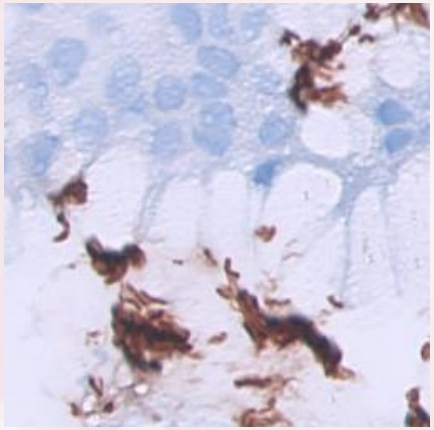


- samples of gastric mucosa
- very large images (120000 x 16000 pixels)

Biopsy analysis is time consuming

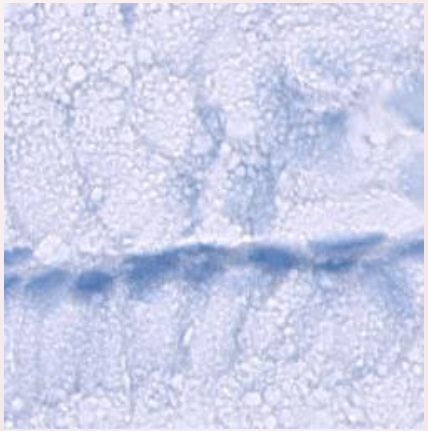
Infected tissue

Immunostained H.Pylori are visible as dark brown spots

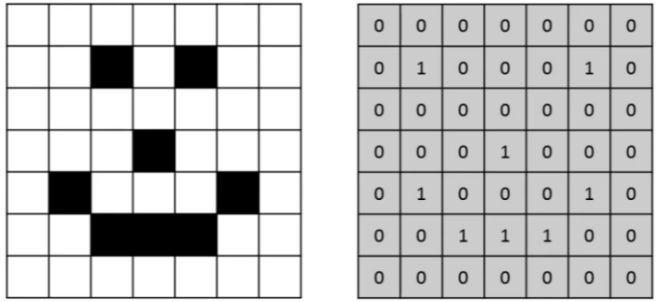


Healthy tissue

Tissue is free of H.Pylori

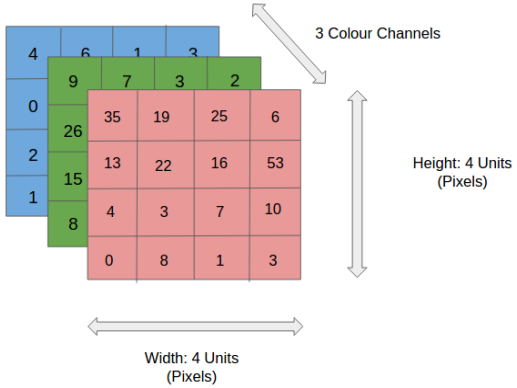


Pictures are just grids of numbers



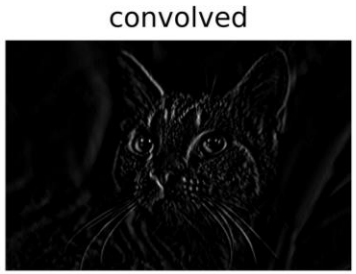
0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Pictures is grid of pixels. Each pixel contains number, eg for grayscale



Often >1 grid per picture to encode colour

Convolutional operations



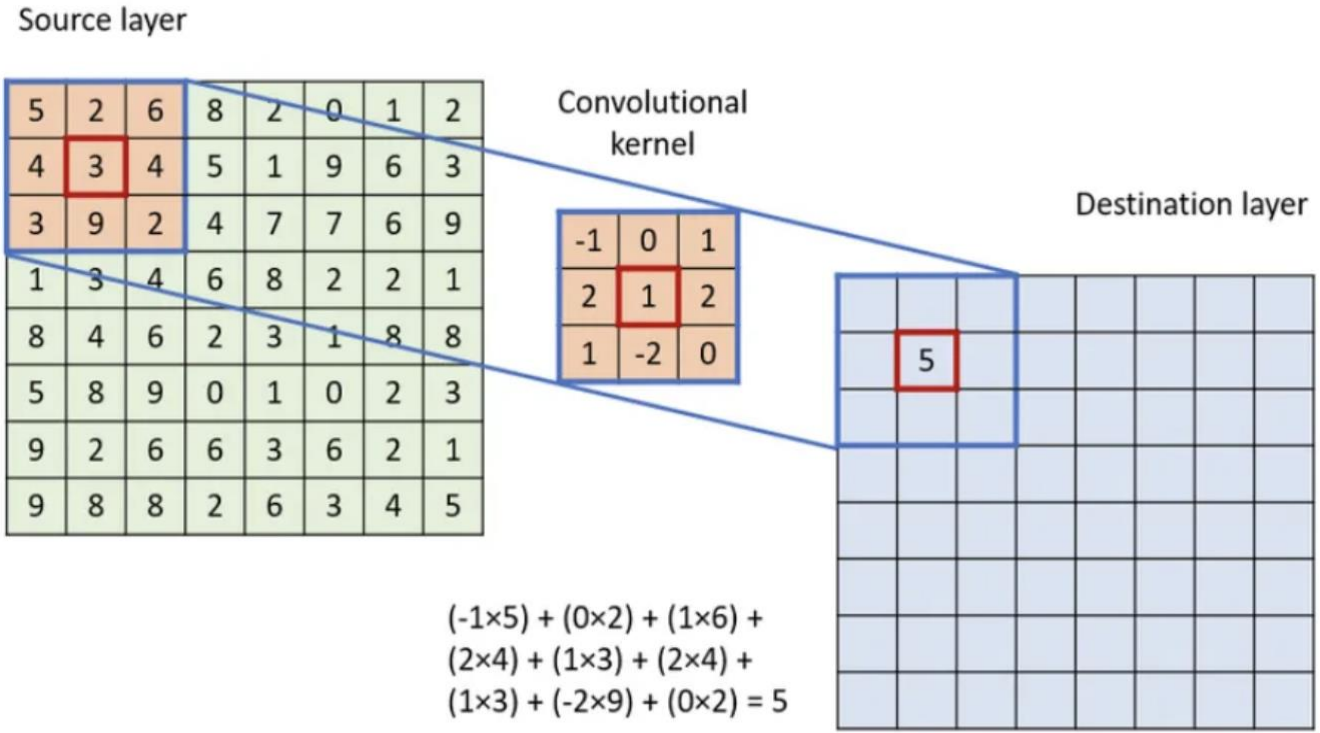
Method:

Apply specialized filters (kernels) to each pixel

Function:

Extract features like edges, textures & shapes.

Mathematics of kernels



Examples of kernels

Original

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

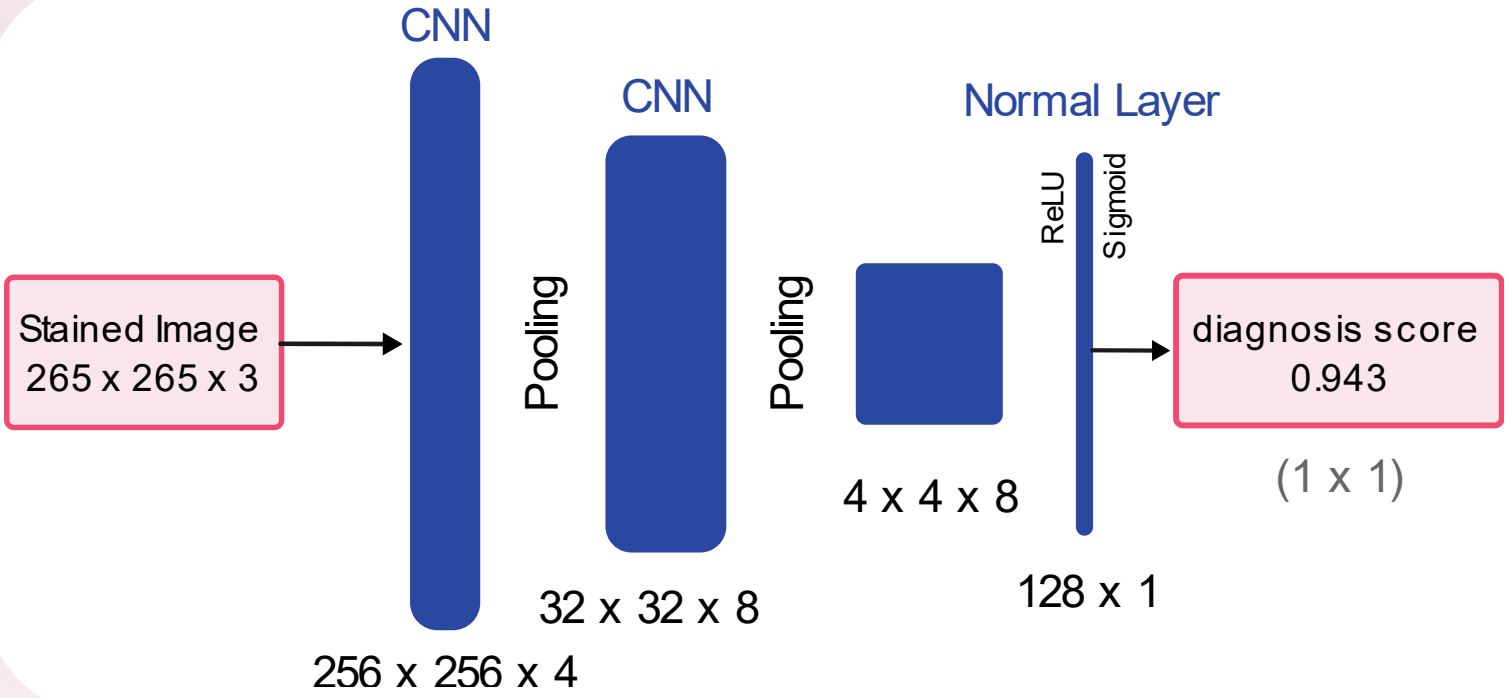
Blur

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

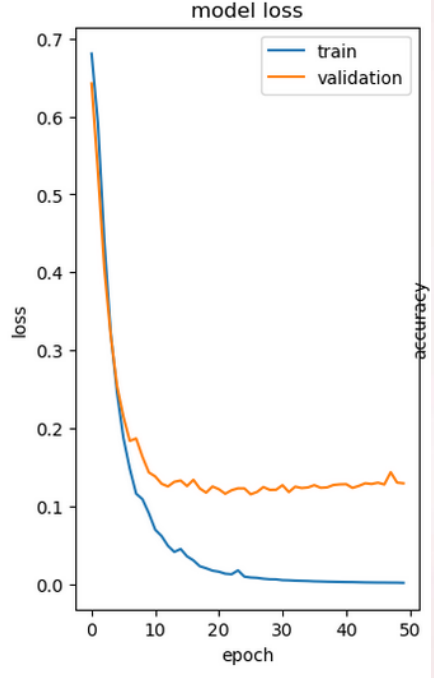
Edge Detection

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Network Design



Training



https://api.wandb.ai/links/reyio-universitat-aut-noma-de-barcelona/z8gdzvj4

Training


val_loss
20241115_190532_103716_4
20241115_190532_103716_3
20241115_190532_103716_2
20241115_190532_103716_1
20241115_190532_103716_0
0.8
0.6
0.4
0.2
50
100
150
200
250
Step

Training Results

- 2400 samples for training
- 600 samples for evaluation

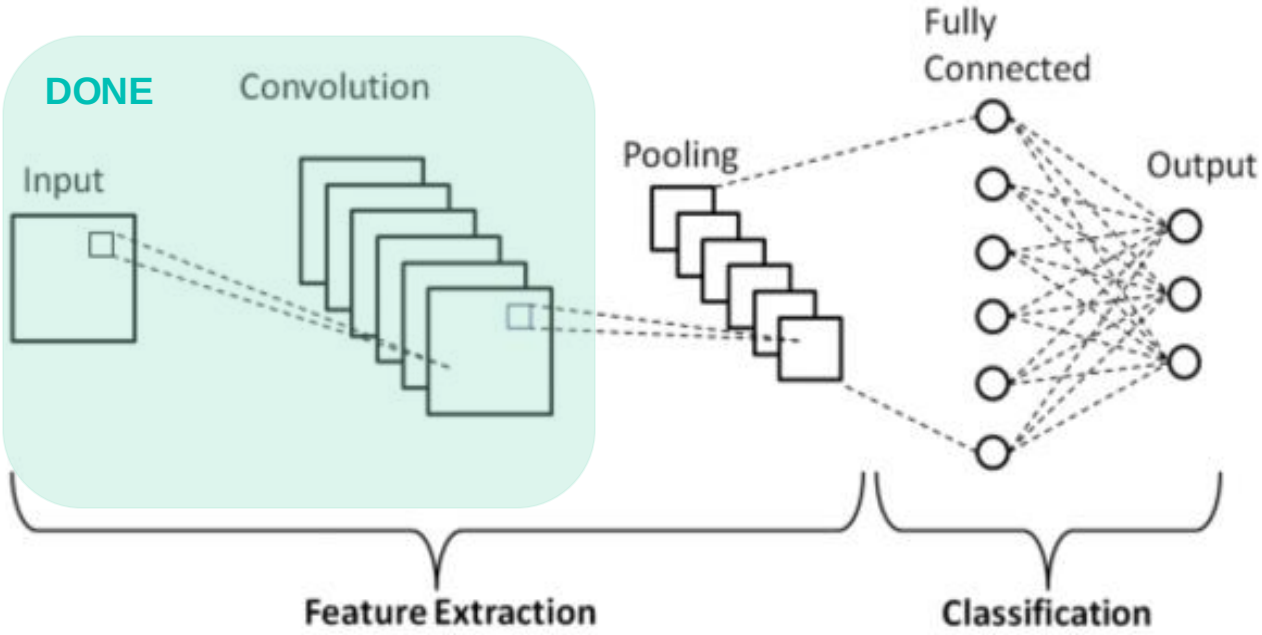
- 120 patients for training
- 30 patients for evaluation

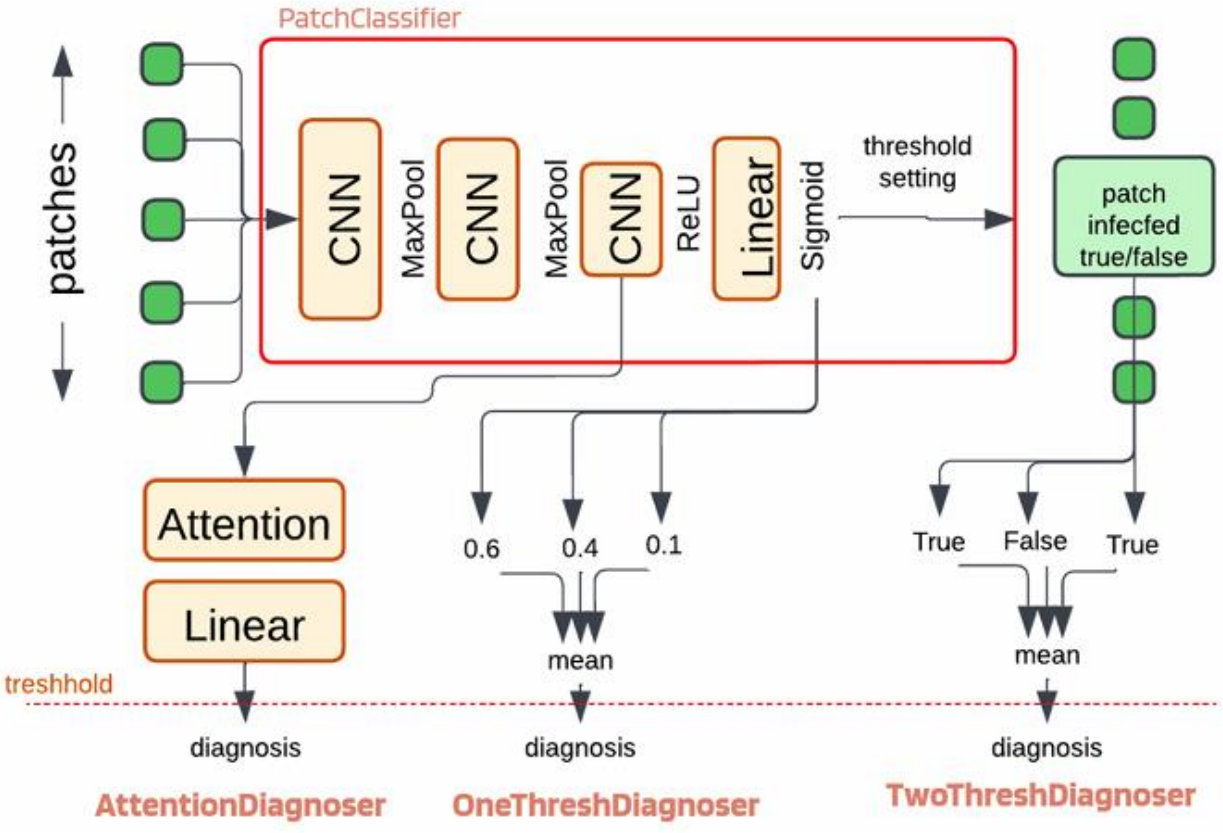
[eval]	neg diagnosis	pos diagnosis
healthy	**53** % ± 12	**3,5 %** ± 3,5
sick	**2,4 %** ± 1,1	**41 %** ± 10

$$\frac{\text{correct diagnoses}}{\text{all diagnoses}} = \mathbf{94}\ \% \pm 4$$

Thanks for your Attention

Next: dimension reduction via pooling




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

test-Cropped / HoldOut	Attention Diagnoser	OneThresh Diagnoser	TwoThresh Diagnoser	Distribution Diagnoser
Accuracy	90% ± 5% / 84% ± 2%	92% ± 3% / 85% ± 2%	92% ± 2% / 85% ± 1%	91 % ± 3 % / 84 % ± 2 %
Precision	89% ± 8% / 95% ± 3%	92% ± 6% / 93% ± 3%	93% ± 7% / 94% ± 3%	93 % ± 8 % / 95 % ± 5 %
Positive Recall	92% ± 5% / 73% ± 4%	92% ± 4% / 76% ± 3%	91% ± 4% / 75% ± 3%	88 % ± 5 % / 73 % ± 3 %
Negative Recall	87% ± 11% / 96% ± 3%	92% ± 6% / 94% ± 3%	94% ± 7% / 95% ± 3%	94 % ± 7 % / 96 % ± 4 %
F1	90% ± 5% / 82% ± 2%	92% ± 3% / 83% ± 2%	92% ± 2% / 84% ± 2%	90 % ± 5 % 82 % ± 2 %