



Malaria

Weronika Orska, Cecilia Gallo & María González del Valle



Disease detection

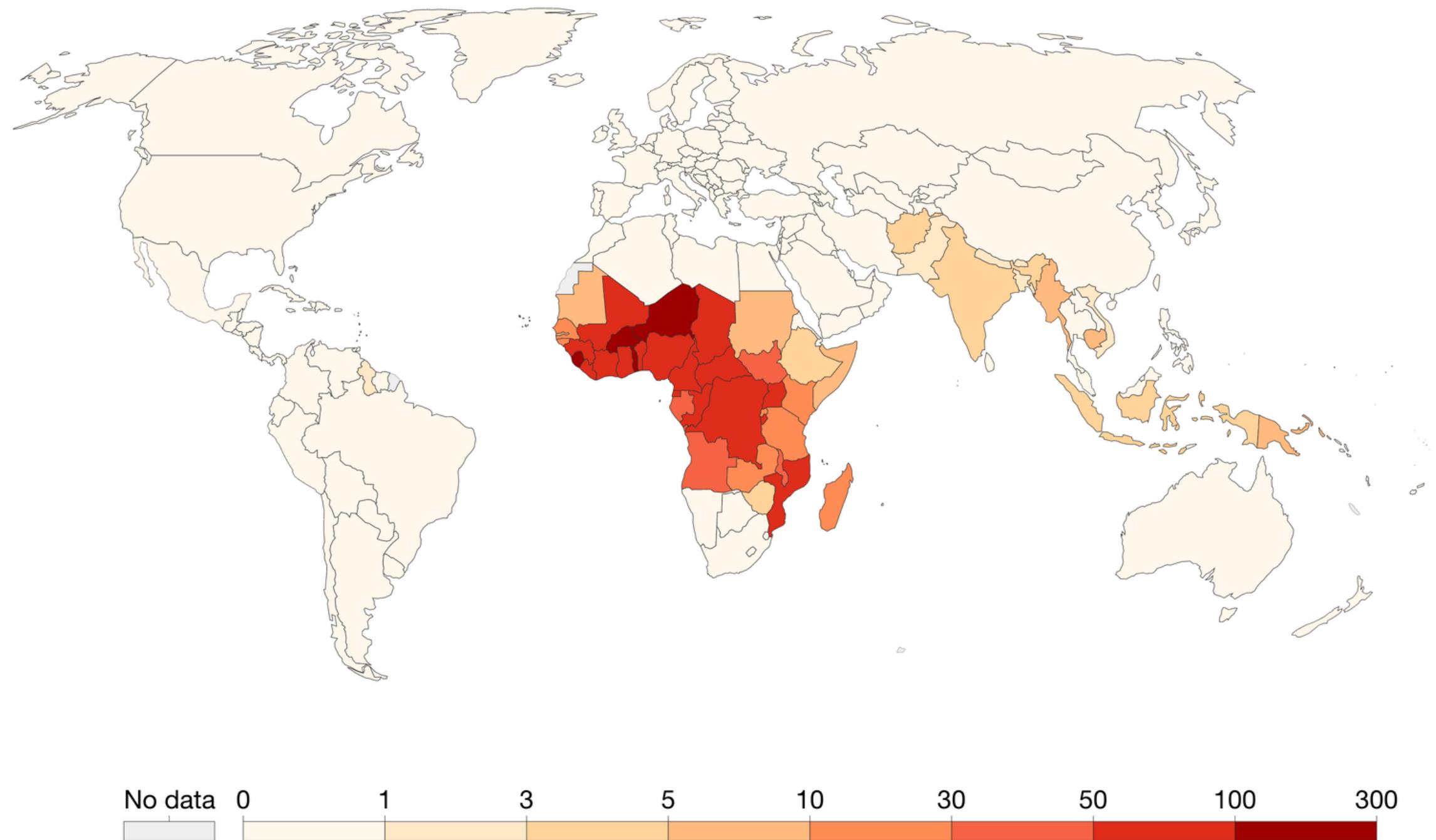


Why are we concerned?

Death rate from malaria

The number of deaths from malaria per 100,000 people.

Our World
in Data



Source: Institute for Health Metrics and Evaluation (IHME)

Note: To allow comparisons between countries and over time this metric is age-standardized.

OurWorldInData.org/malaria • CC BY

**2 million
people die every year**

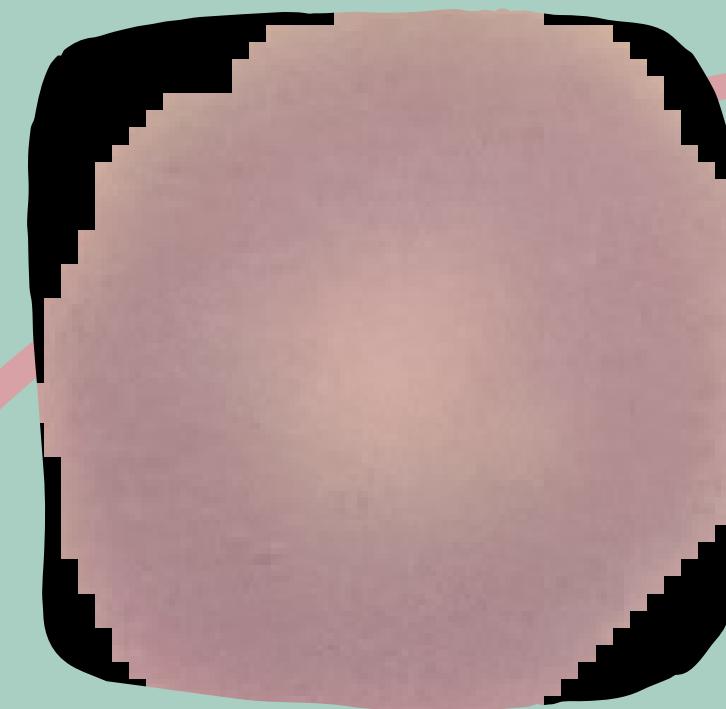
**300-500
million
cases annually**

Malaria Dataset

thin-blood smear slide images



Parasitized



Uninfected

Objectives

 Improve traditional mechanism to detect malaria

 Develop an automated system

 Achieve high accuracies

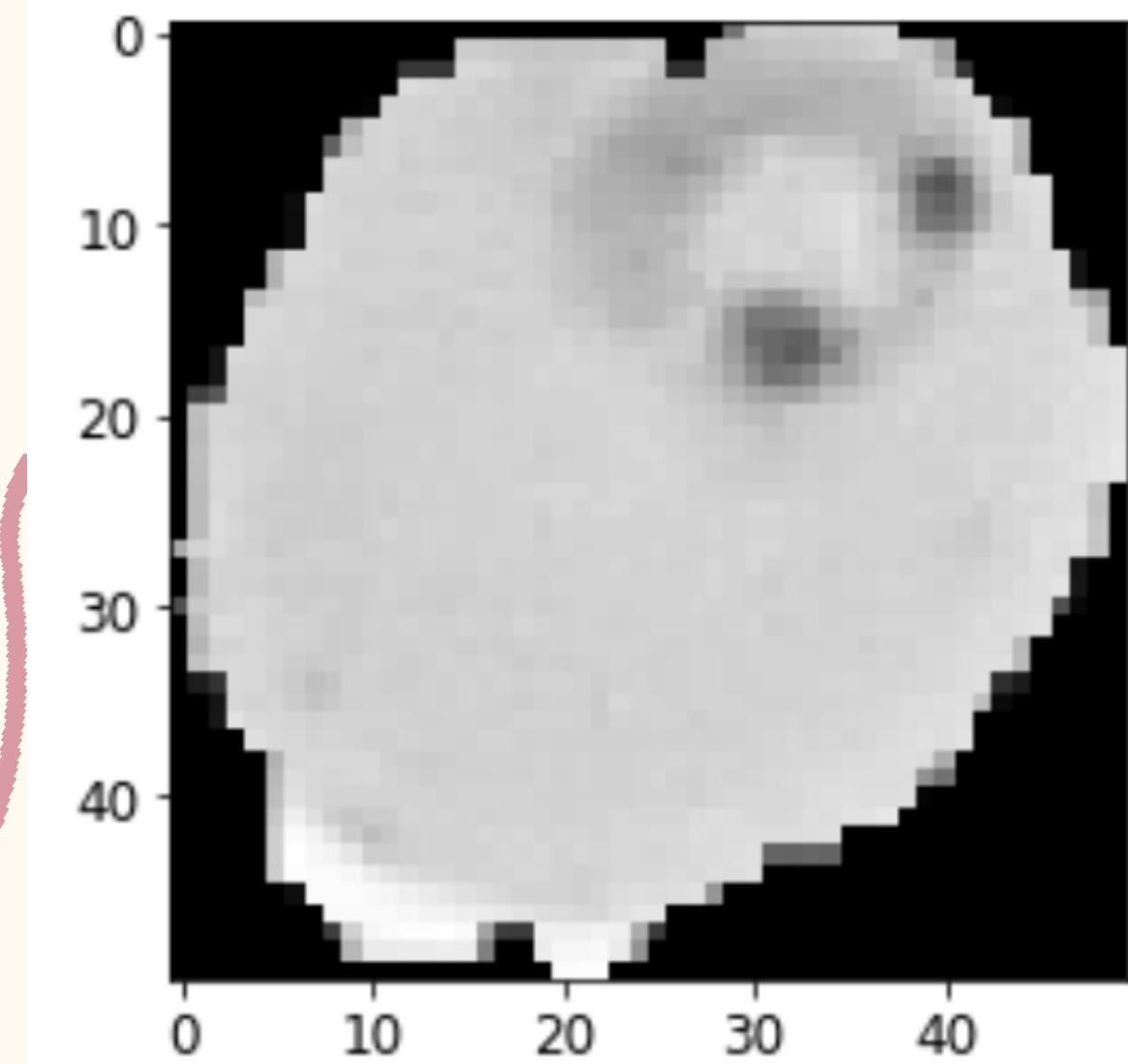
Our model

First steps

- Converted images to grey scale
- Made the image shapes the same --> 50x50
- Reduced definition
- Labelled and shuffled our data
- Normalized it

Implementing the model

- Libraries used -> Keras and Tensorflow
- CNN with 3 layers
 - > Loss = Binary-cross entropy
 - > Batch size = 32



Model Evaluation

- Learning Curve

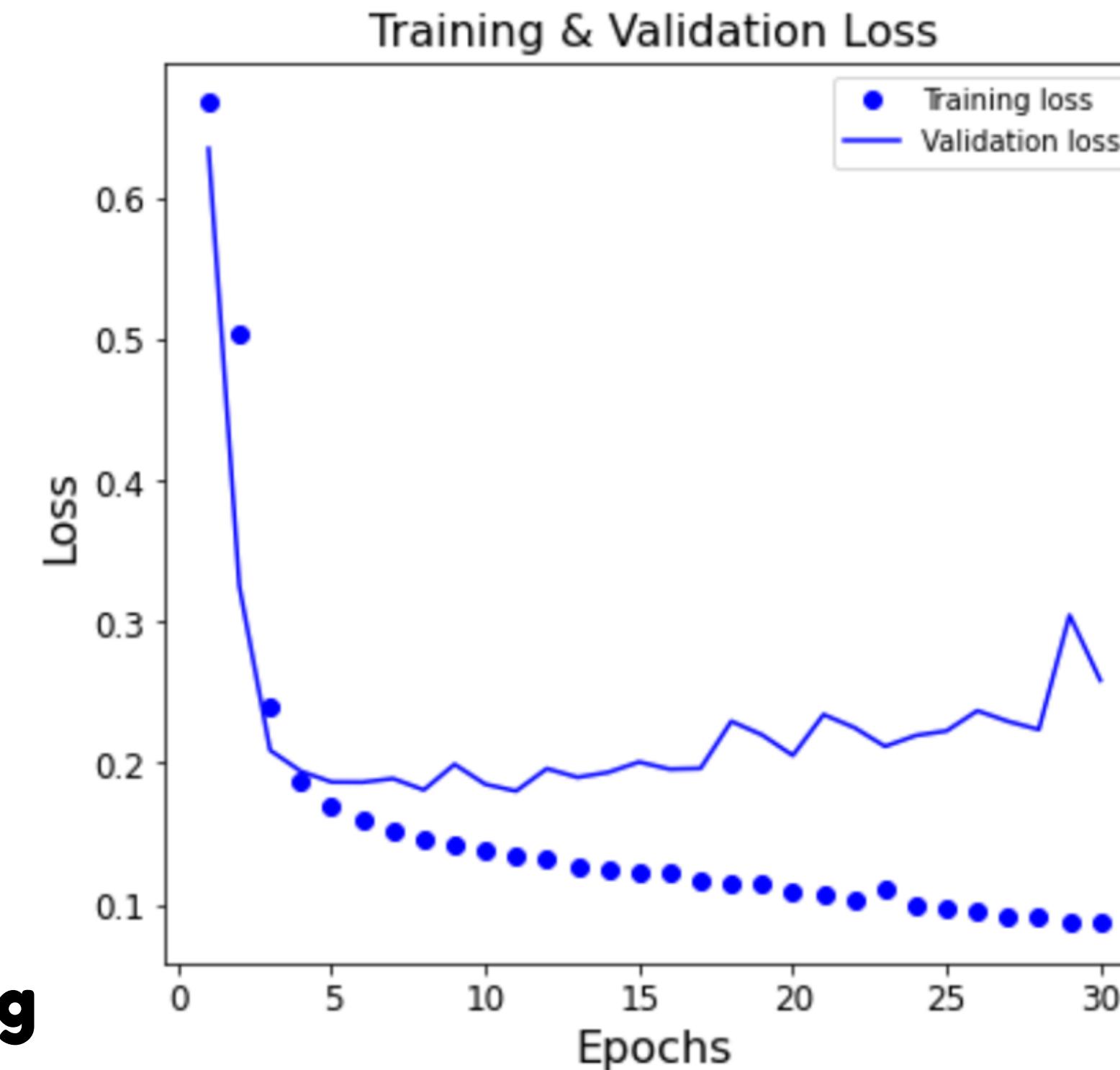
- Accuracy Curve

- Classification Matrix

Learning Curve

- The plot of training loss continues to decrease with experience.
- The plot of validation loss decreases to a point and begins increasing again

...overfitting



Evaluation Metric – Accuracy

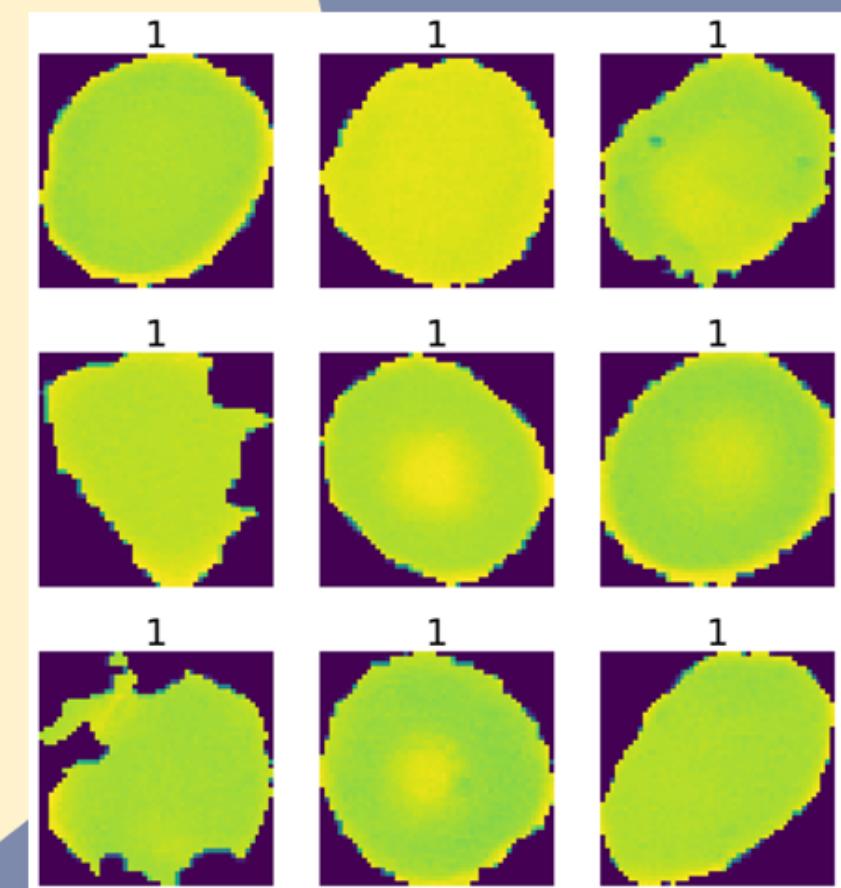
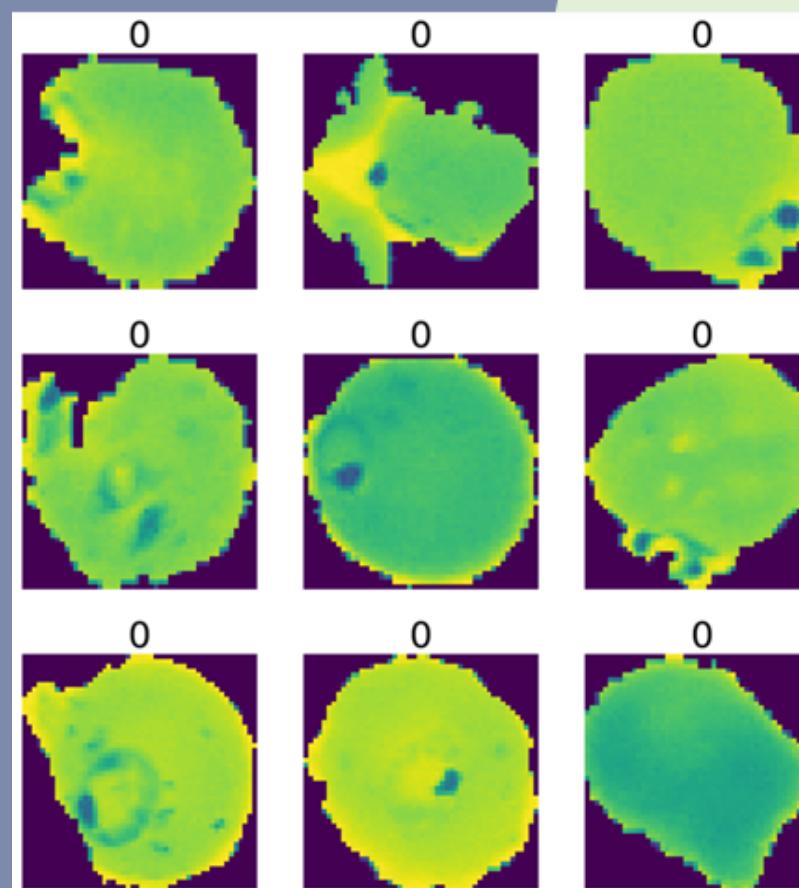
Parasitized

13779

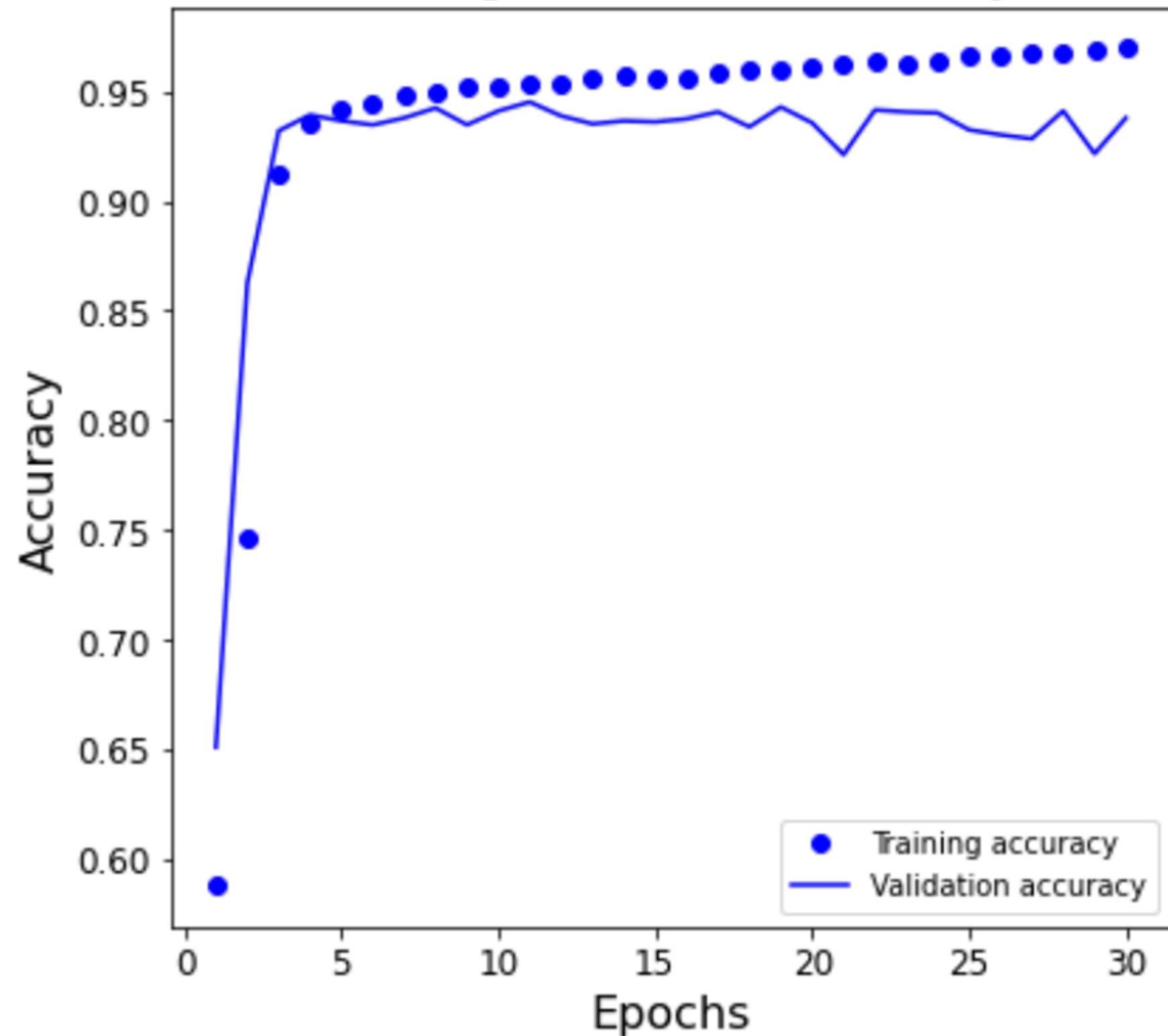
Uninfected

13779

y
(target)



Training & Validation Accuracy



Training accuracy is too high whereas the validation accuracy is less.
...overfitting

Val Accuracy evolution

- 5 epoch: 0.9365
- 10 epoch: 0.9410
- 15 epoch: 0.9361
- 20 epoch: 0.9356
- 25 epoch: 0.9324
- 30 epoch: 0.9379

Classification Matrix

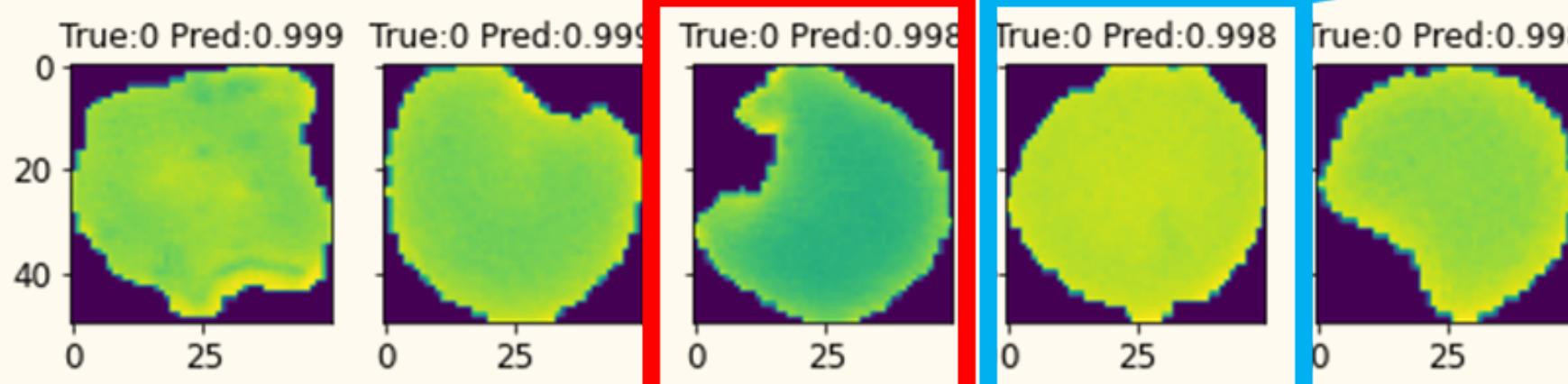
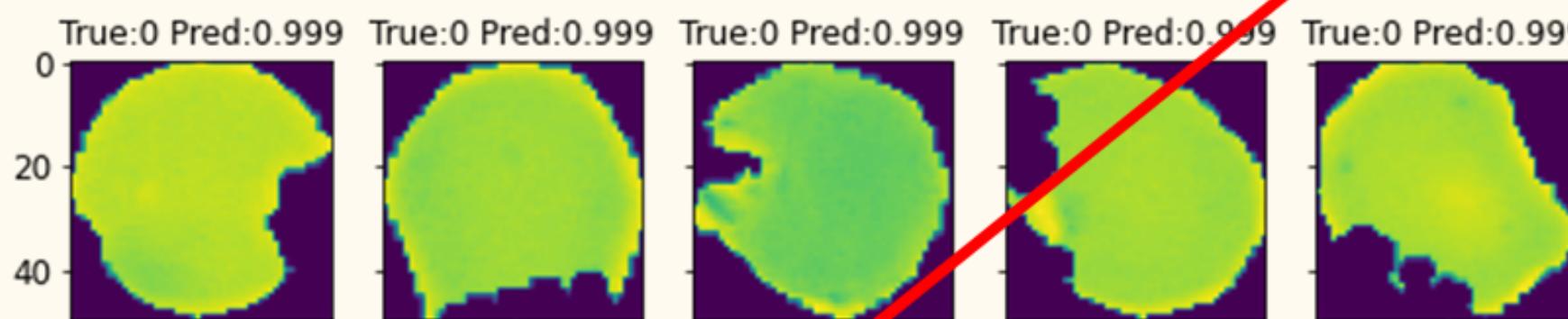
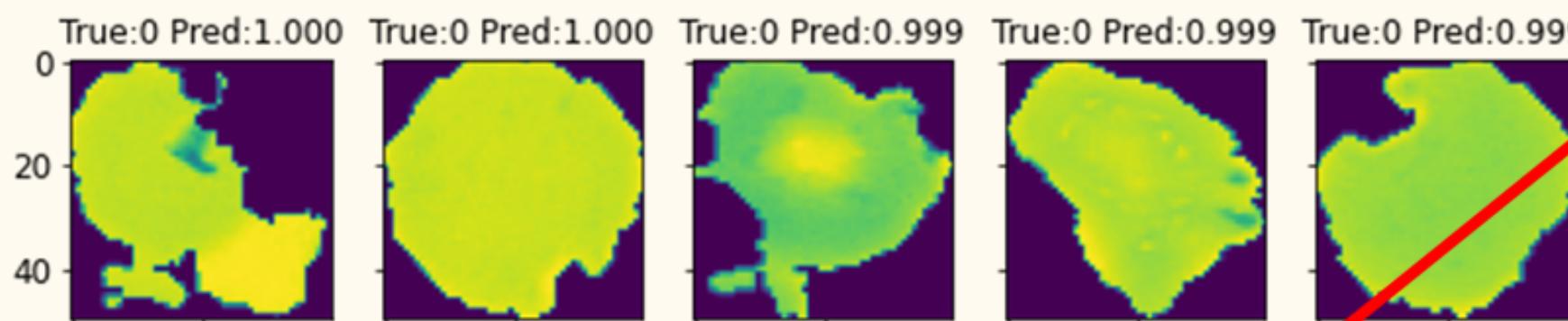
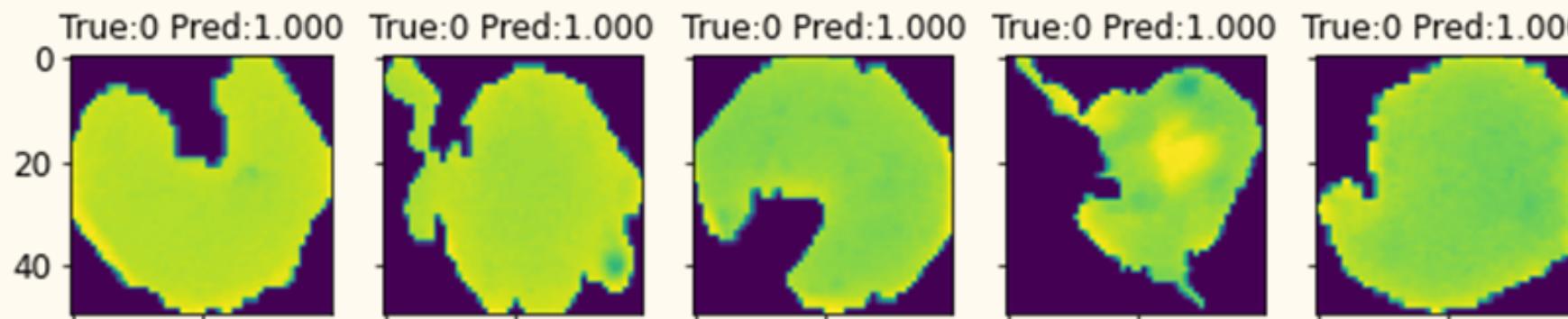
		Predicted	
		0	1
		0	1
Actual	0	2609	168
	1	177	2558

	precision	recall
0	0.94	0.94
1	0.94	0.94
accuracy		
macro avg	0.94	0.94
weighted avg	0.94	0.94

Test set Size = 5512

Number of mispredictions = 345

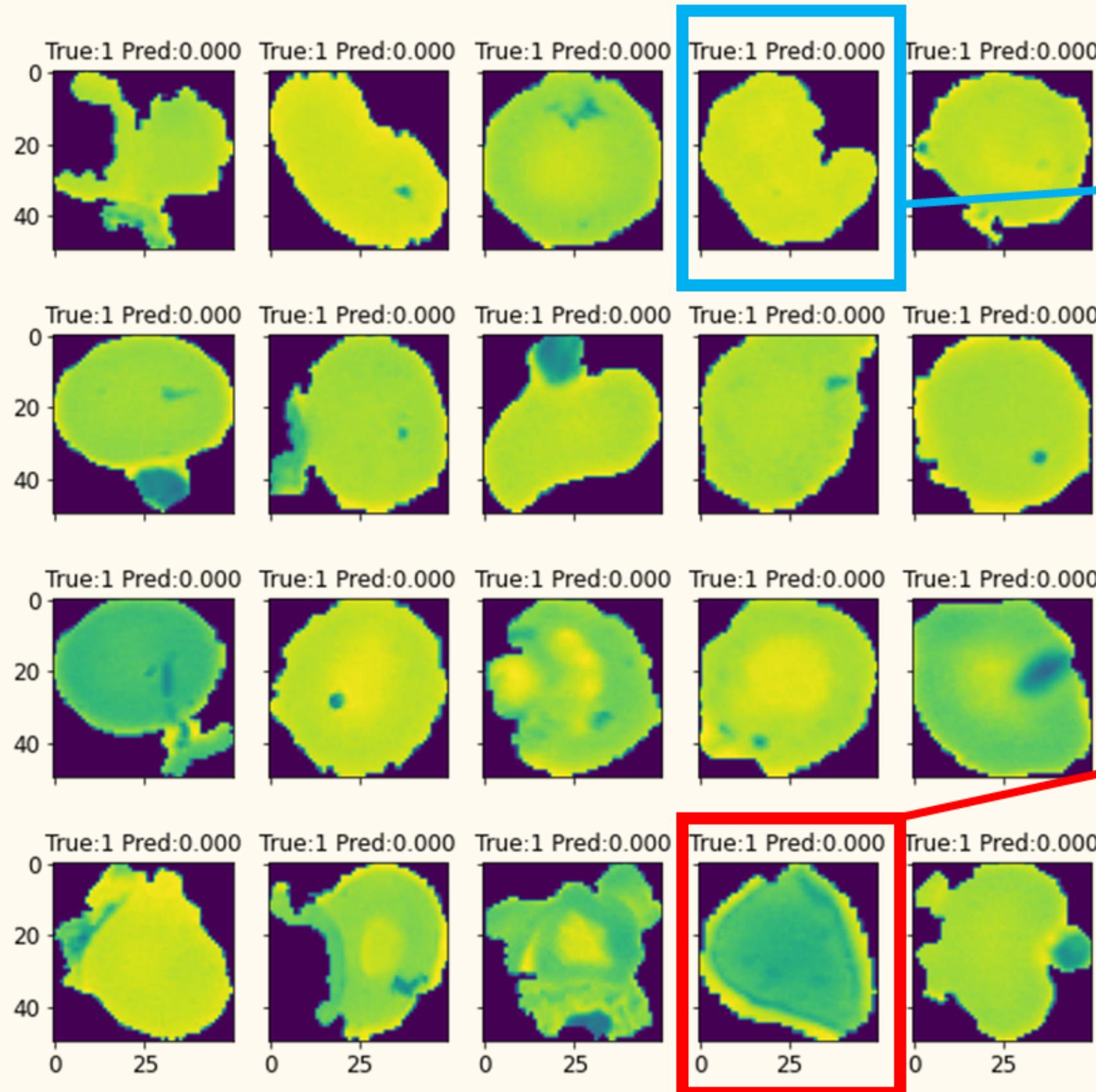
False Positives, 168



**At glance → Parasitized
In reality → Parasitized
Model Prediction → Uninfected**

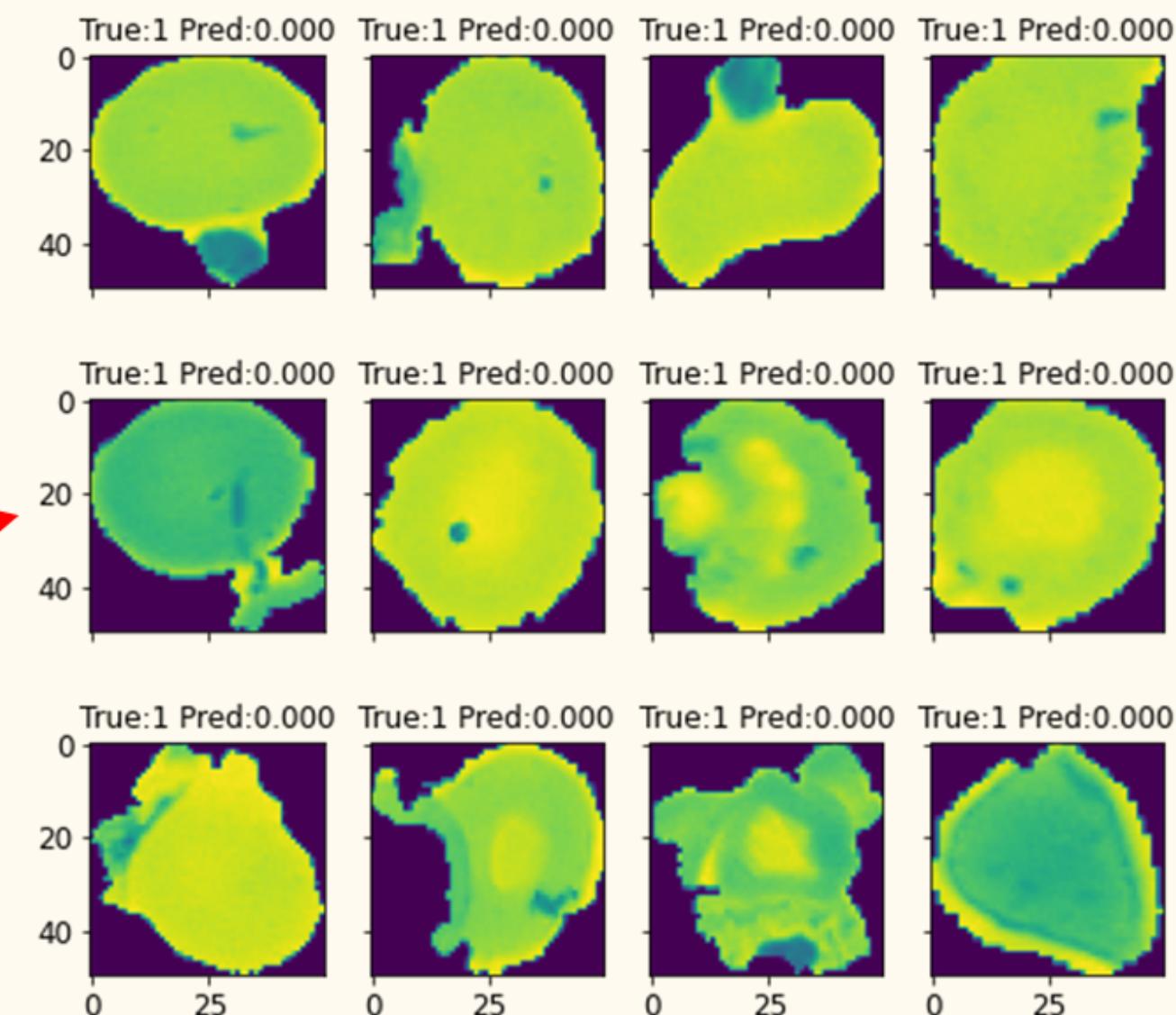
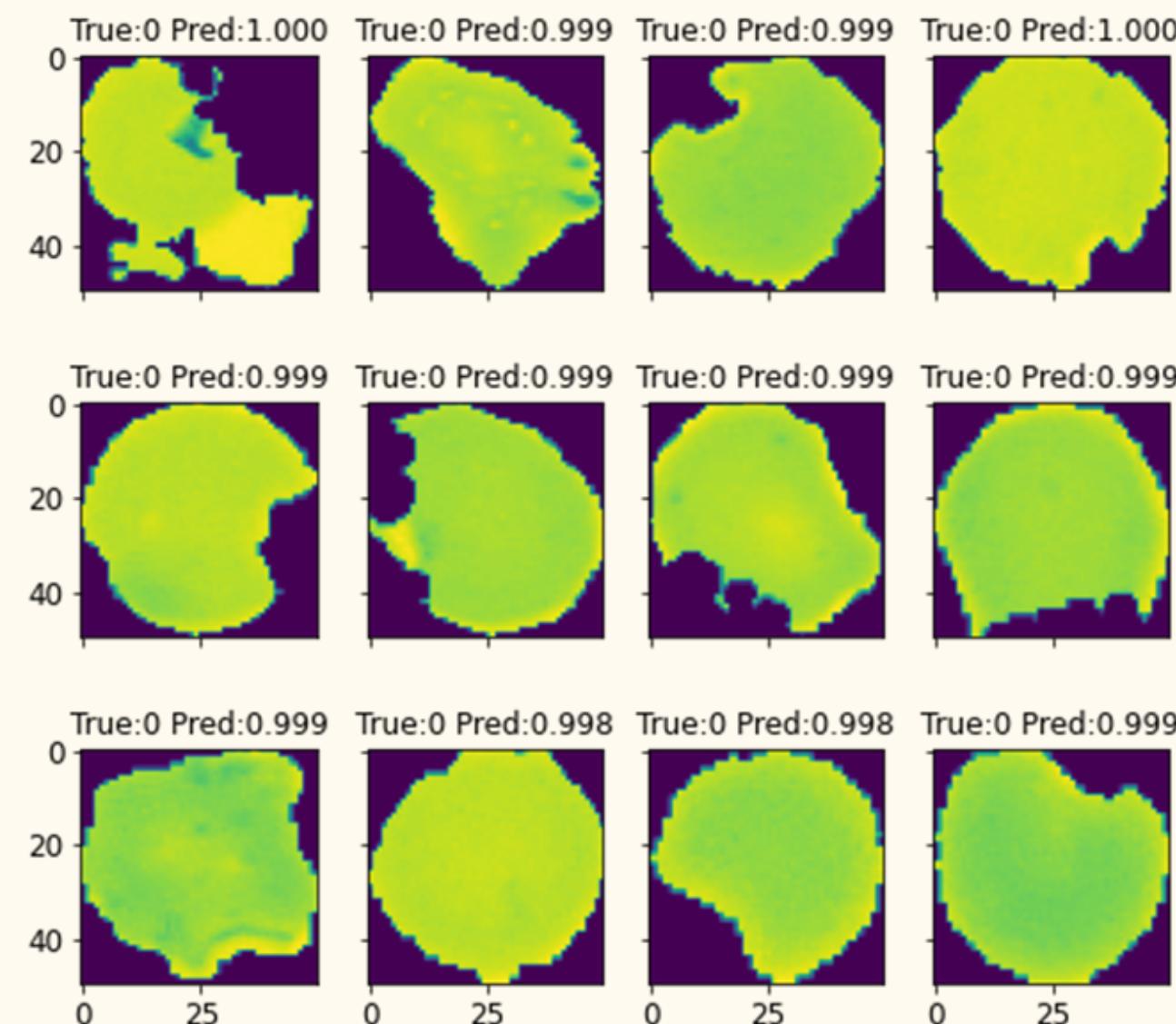
**At glance → Uninfected
In reality → Parasitized
Model Prediction → Uninfected**

False Negatives, 177



Considerations

MORE PROBLEMATIC!



IS THE DATA RELIABLE?

Business Implementation



Thank you!

Any questions?