MALARIA DETECTION

A Deep Learning Model

ERMA C. A.

AGENDA

- Background

- Problem Summary

- Solution Design

- Model Comparison

- Model Selection

- Recommendations for Implementation

- Questions

BACKGROUND

- Over **400,000** people die of malaria each year worldwide

- Estimated that 2/3 of the deaths are **children under five**

- Leading cause of death and disease in developing countries

- Bloodborne disease caused by plasmodium parasites that infect red blood cells (RBCs)

- **Curable** if diagnosed and treated promptly

- Delays in diagnosis are the leading cause of death in patients

- Microscopic examination is the "gold standard" for diagnosis

PROBLEM SUMMARY

- **What is the problem?**

  - Malaria diagnosis is time-consuming process

  - Diagnosis accuracy is highly dependent on experience of lab personnel

- **How are we going to help solve the problem?**

  - Build an efficient computer vision model to examine RBC image

  - Model will identify the image

    - Infected with malaria and classifying as parasitized

    - Not infected and classifying as uninfected
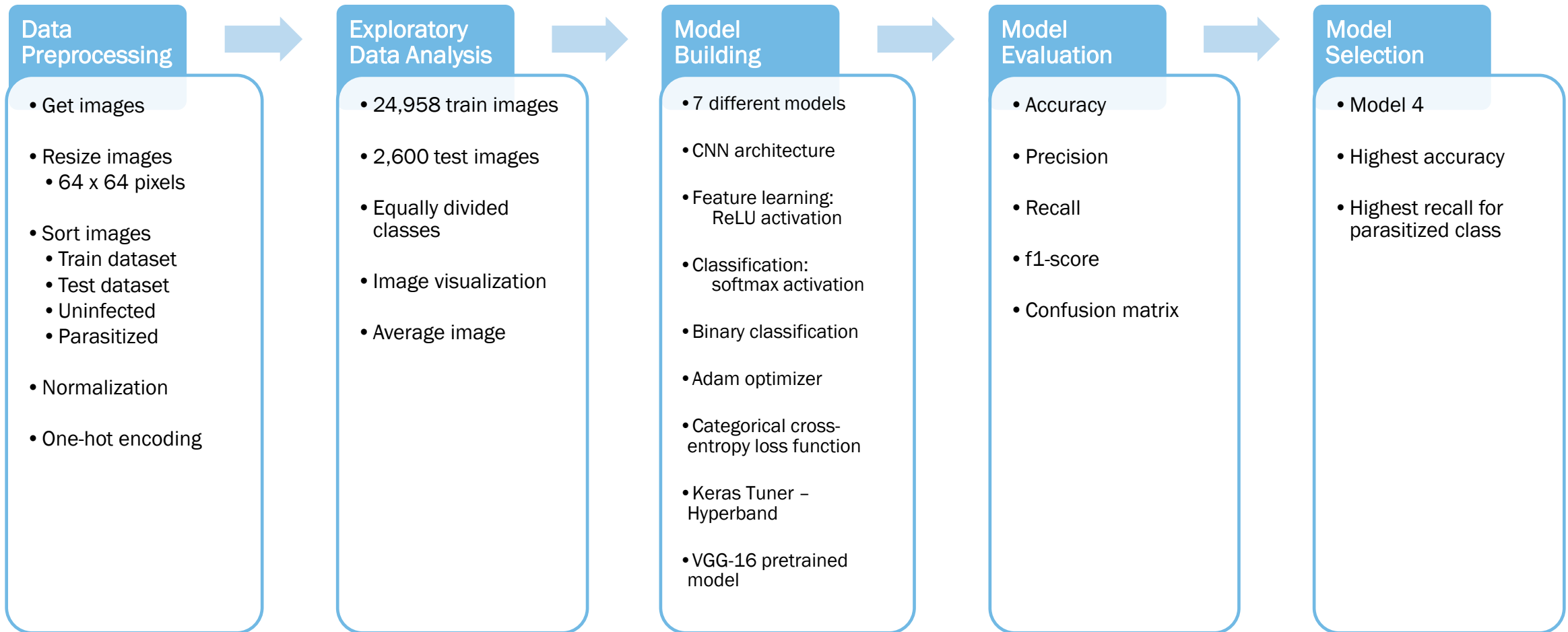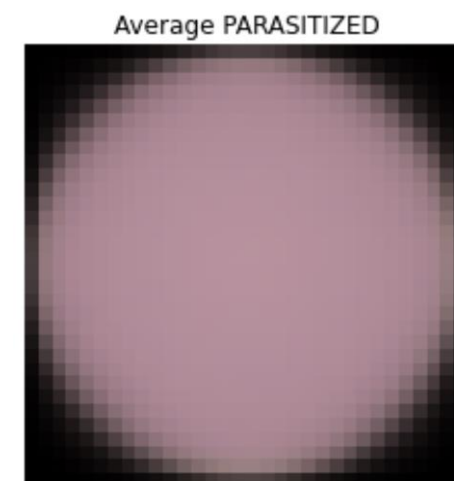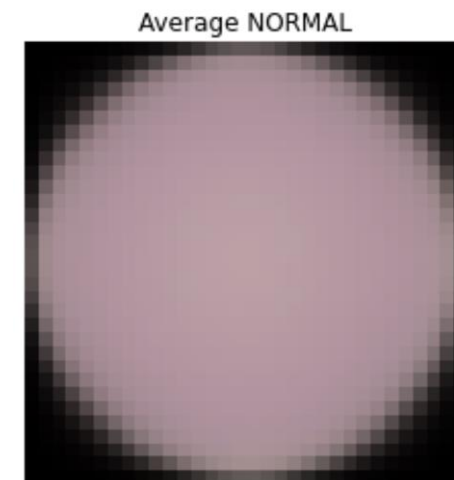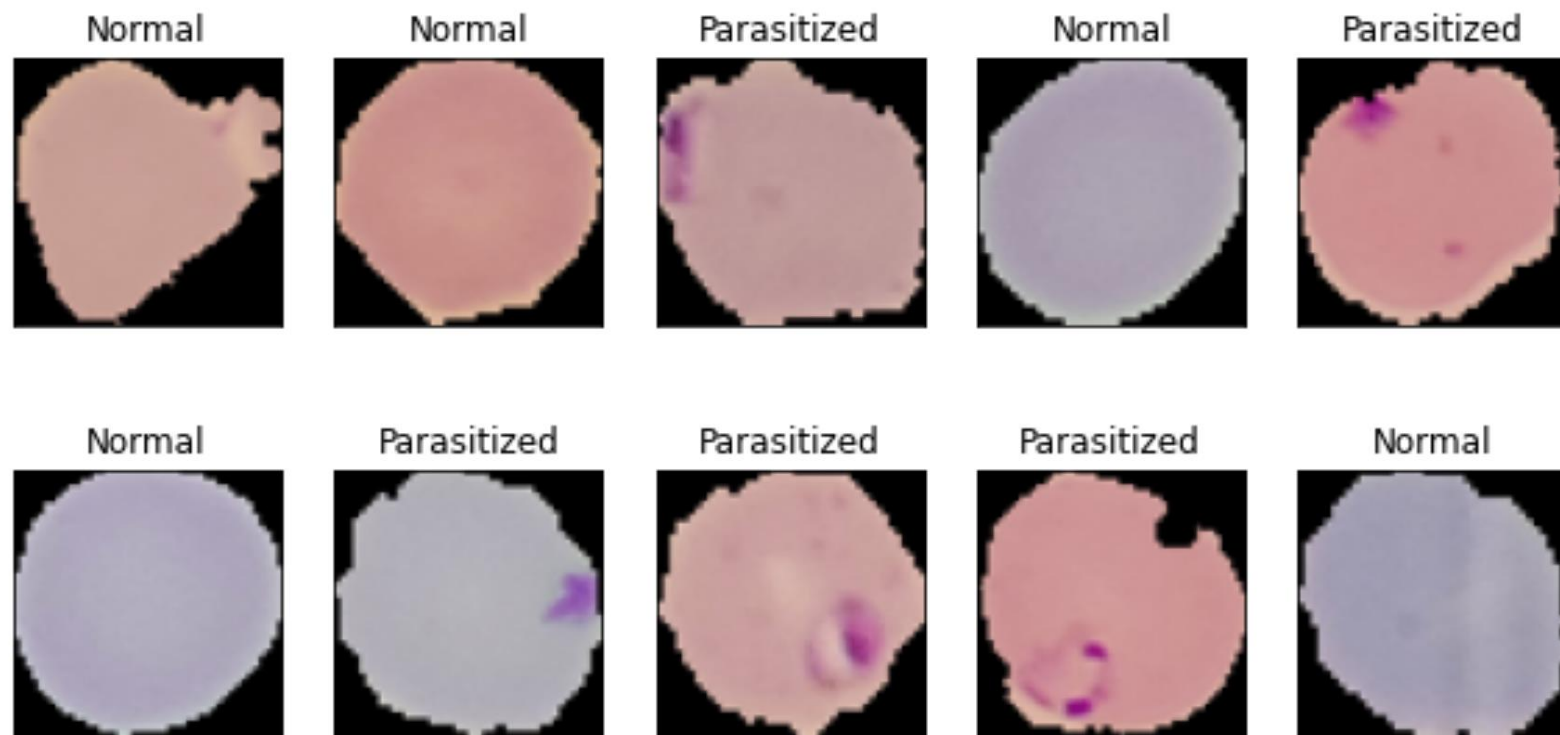
SOLUTION DESIGN

Data Preprocessing

- Get images
- Resize images
  - 64 x 64 pixels
- Sort images
  - Train dataset
  - Test dataset
  - Uninfected
  - Parasitized
- Normalization
- One-hot encoding

Exploratory Data Analysis

- 24,958 train images
- 2,600 test images
- Equally divided classes
- Image visualization
- Average image

Model Building

- 7 different models
- CNN architecture
- Feature learning: ReLU activation
- Classification: softmax activation
- Binary classification
- Adam optimizer
- Categorical cross-entropy loss function
- Keras Tuner – Hyperband
- VGG-16 pretrained model

Model Evaluation

- Accuracy
- Precision
- Recall
- f1-score
- Confusion matrix

Model Selection

- Model 4
- Highest accuracy
- Highest recall for parasitized class

IMAGE VISUALIZATION

MODEL COMPARISON: ACCURACY

Train Accuracy

Validation Accuracy

MODEL COMPARISON: LOSS

Train Loss
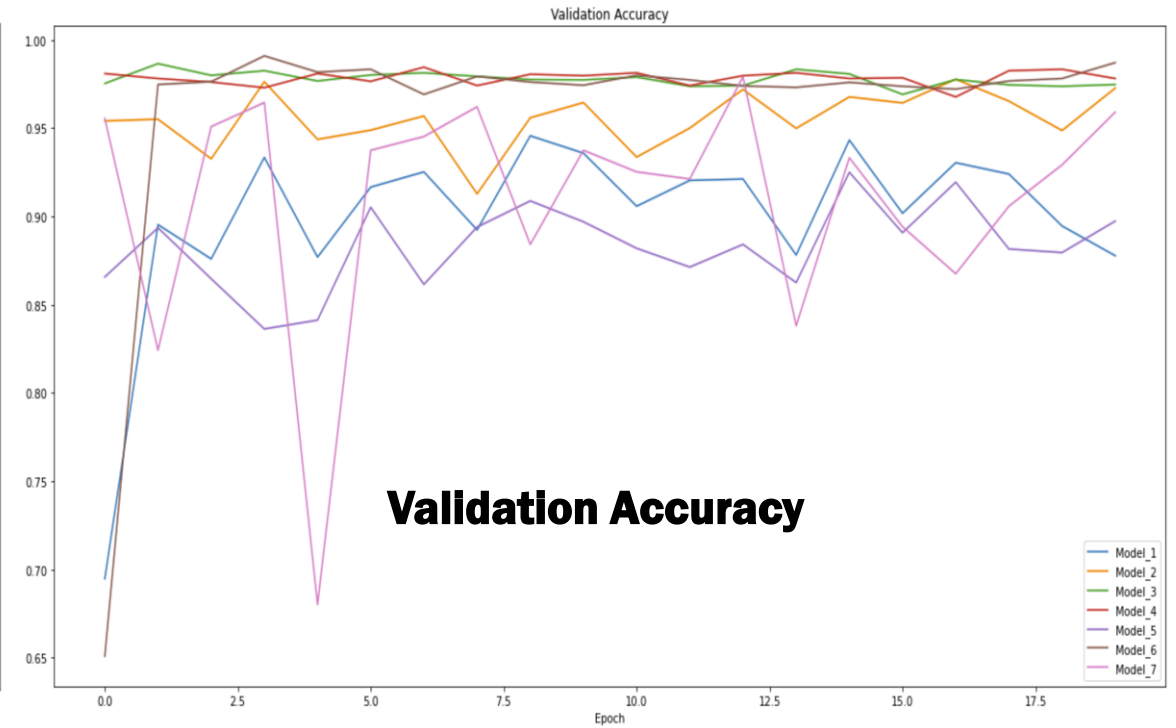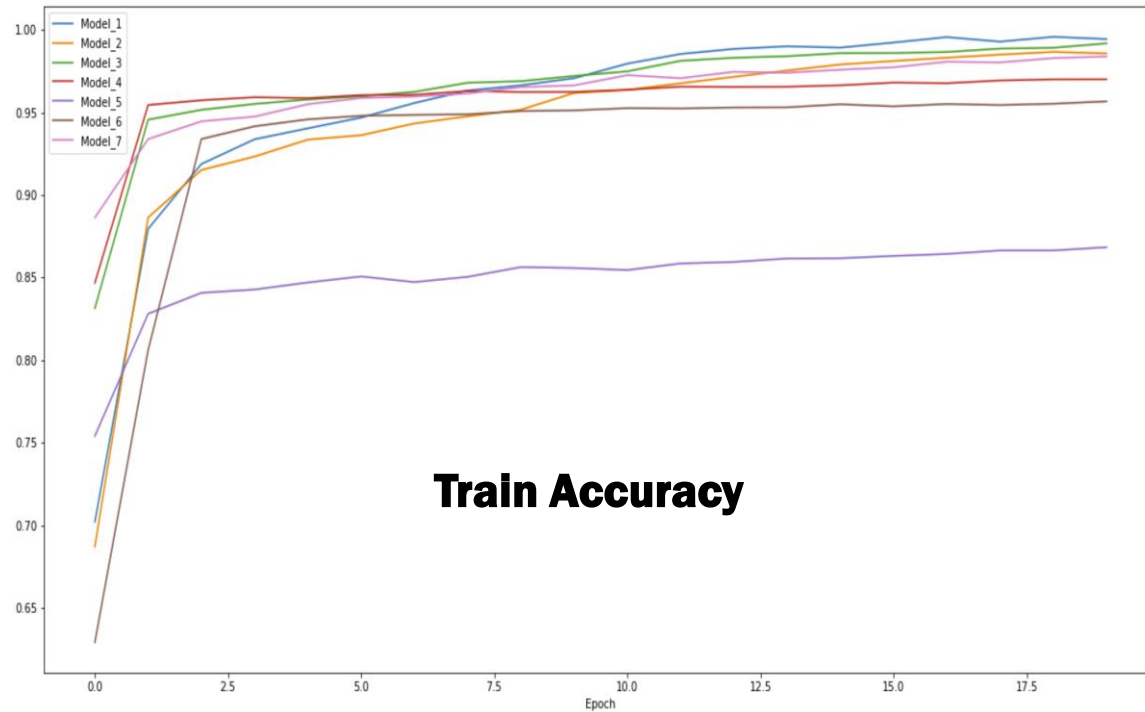
Validation Loss
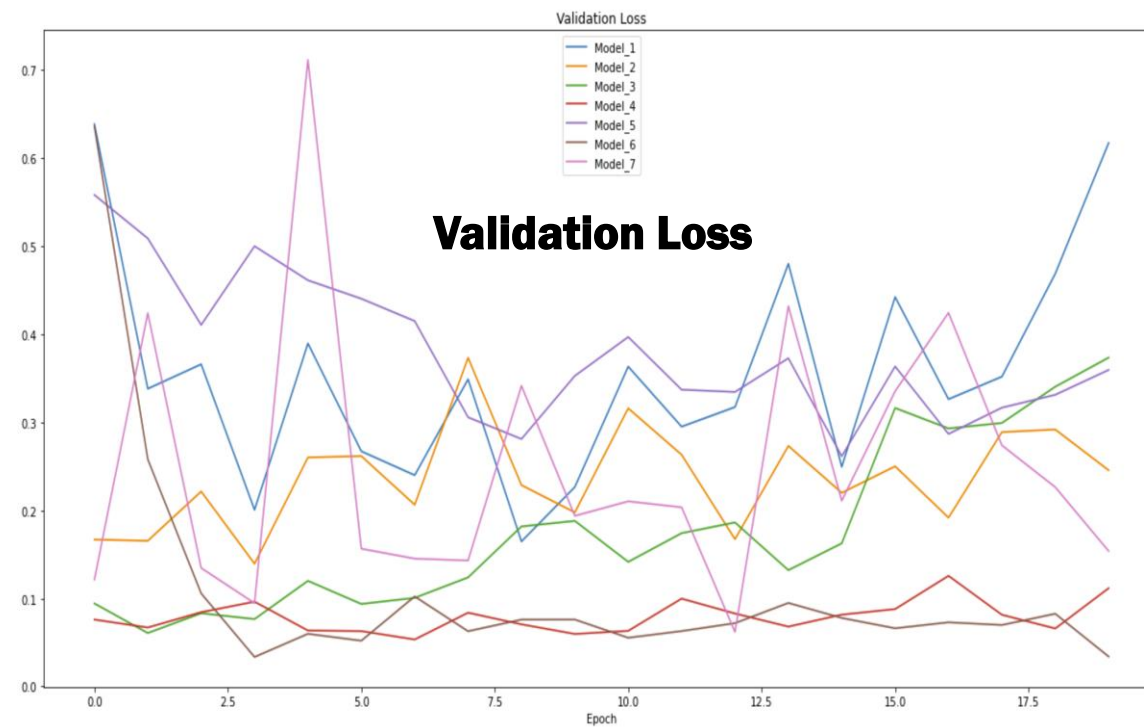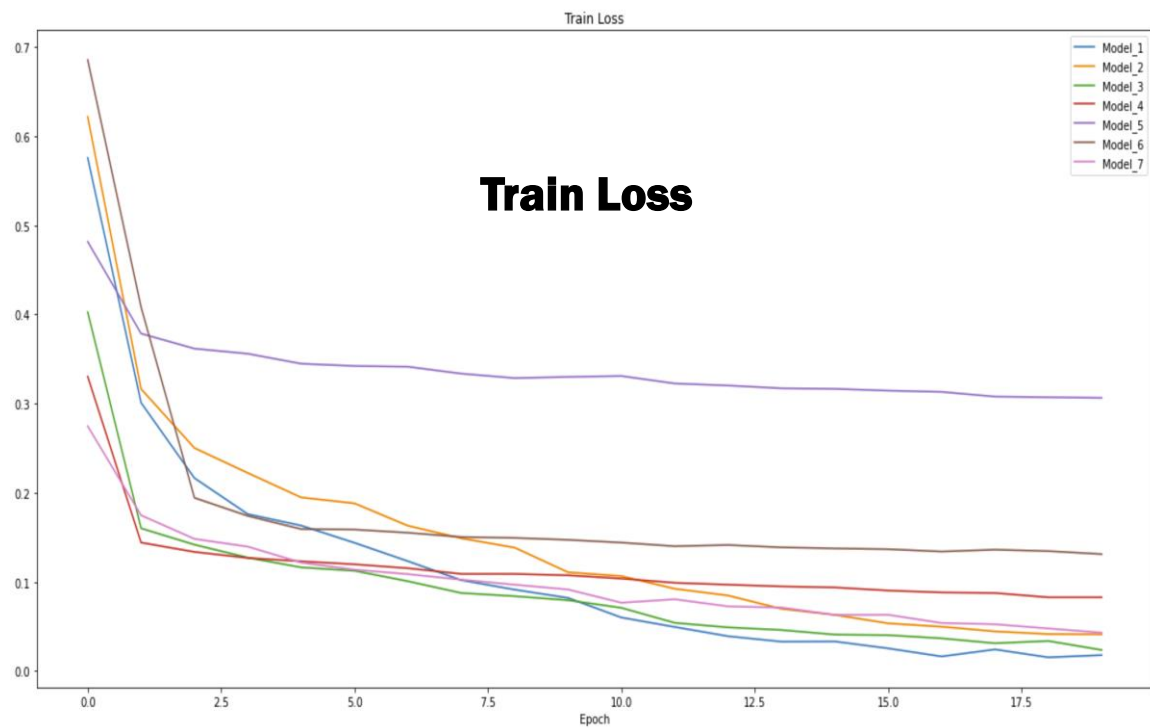
MODEL COMPARISON

Class			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
		Accuracy	0.91	0.93	0.94	0.95	0.88	0.94	0.95
Class	**Uninfected**	**Precision**	0.94	0.91	0.92	0.94	0.86	0.91	0.93
		Recall	0.87	0.97	0.97	0.97	0.9	0.98	0.96
		f1-score	0.9	0.94	0.95	0.96	0.88	0.94	0.95
	Parasitized	**Precision**	0.88	0.97	0.97	0.97	0.89	0.98	0.96
		Recall	0.94	0.9	0.92	0.94	0.85	0.9	0.93
		f1-score	0.91	0.93	0.94	0.95	0.87	0.94	0.95

MODEL SELECTION: MODEL 4

- **Feature Learning:** 4 CNN layers, LeakyReLU activation, dropout, max pooling

- **Classification:** 2 classes, softmax activation

- Adam optimizer

- Categorical cross-entropy loss function

- **Accuracy:** 95%

- **True positive rate (TPR):** 94%

- **False negative rate (FNR):** 6%
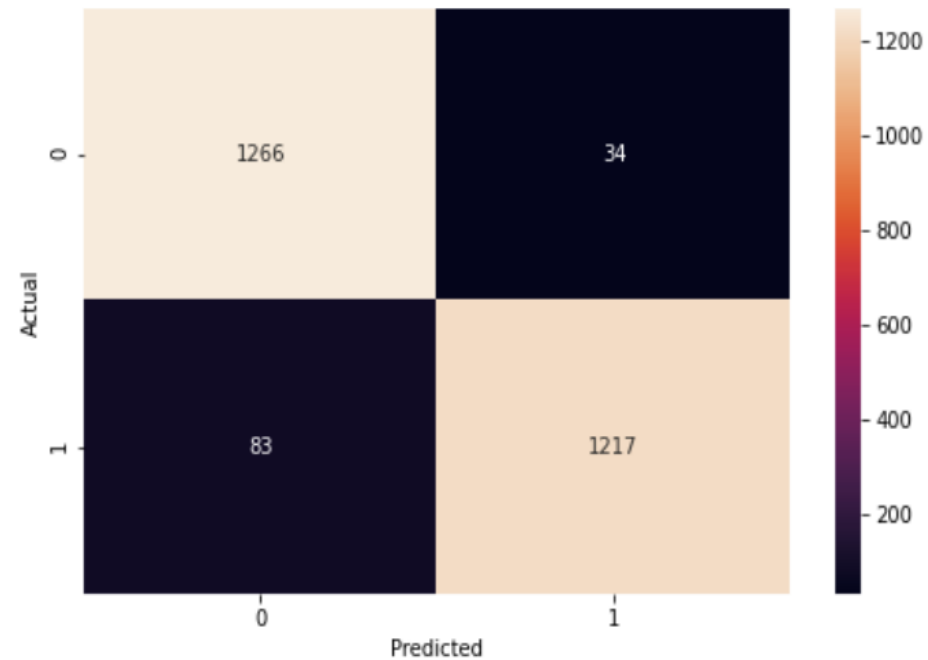
actual: Parasitized
predicted: Normal
probability: 1.0
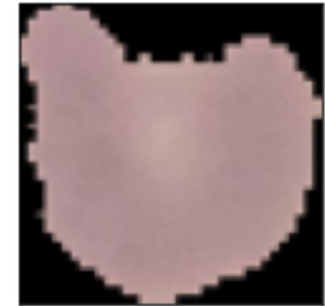
actual: Parasitized
predicted: Normal
probability: 1.0

RECOMMENDATIONS FOR IMPLEMENTATION

- Model 4 is best performing model for computer vision system

- **Next steps:** additional computing power, data augmentation, images resized with more pixels

- **Current Model Limitations:** color images, single RBC image, 64 x 64 pixels

- **Key Benefits:**

  - ↓ diagnosis time & ↑ accuracy → reduction of the number of deaths

  - Cost avoidance regarding lab personnel → U.S. Lab Technician $36,710 vs Pathologist $283,900

- **Risks to consider:**

  - False negative → possible **DEATH**

  - False positive → unnecessary treatment & cost

THANK YOU