# Malaria Detection

## Deep Learning

## Final Presentation

# Contents / Agenda

- Business Problem Overview
- Problem Statement
- Data Overview
- Data Preprocessing
- Model Execution Results
- Model Performance Summary
- Conclusion

# Business Problem Overview

## Malaria

- Tropical and Sub-Tropical disease
- Worldwide cases 247 million in 2021.
- Caused by mosquito bites
- Disease doesn’t spread from person to person.
- Symptoms include fever, chills and headaches
- Preventable and Curable, with severe cases leading to death.
- Gold Standard in detection includes observing blood smear under microscope.

## Advantages

- Simple technique
- Gives results quickly.

## Challenges

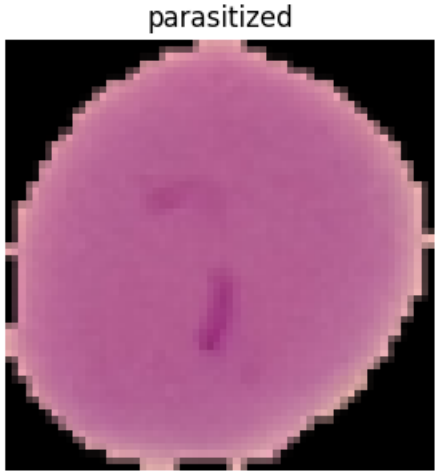
- Observing blood cells requires expertise.
- Needs to be done quickly.

## Problem Statement

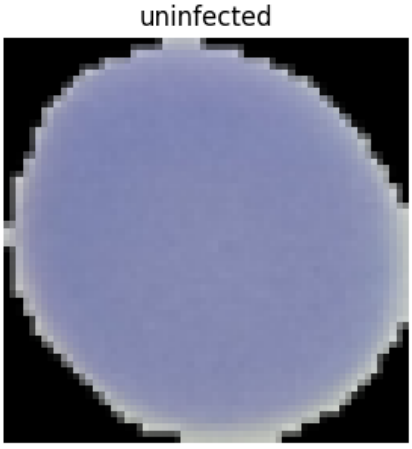
- Can Deep Learning help in the detection of Malaria efficiently and with minimal resources required.

## Data Overview

- **Parasitized Red Blood Cell**




- **Uninfected Red Blood Cell**



## Data Pre-Processing

- Read in the images.
- Standardized to a 64x64 pixel image.
- Images normalized to remove any image capture biases.
- Image encoded to improve model performance.

# Model Performance Summary

- Run 8 different models.
- Two Data Augmentation approaches were employed.
- One Pre-Built model was used (VGG16)
- Run few models as a Binary Classification model.

## Elements of the Models

- Used 32 and 64 nodes
- Had between 13 and 21 layers
- Activation functions used were Relu, Leakyrelu, Softmax and Sigmoid.
- Loss Functions used were Binary Crossentropy and Categorical Crossentropy
- Optimizers used was Adam

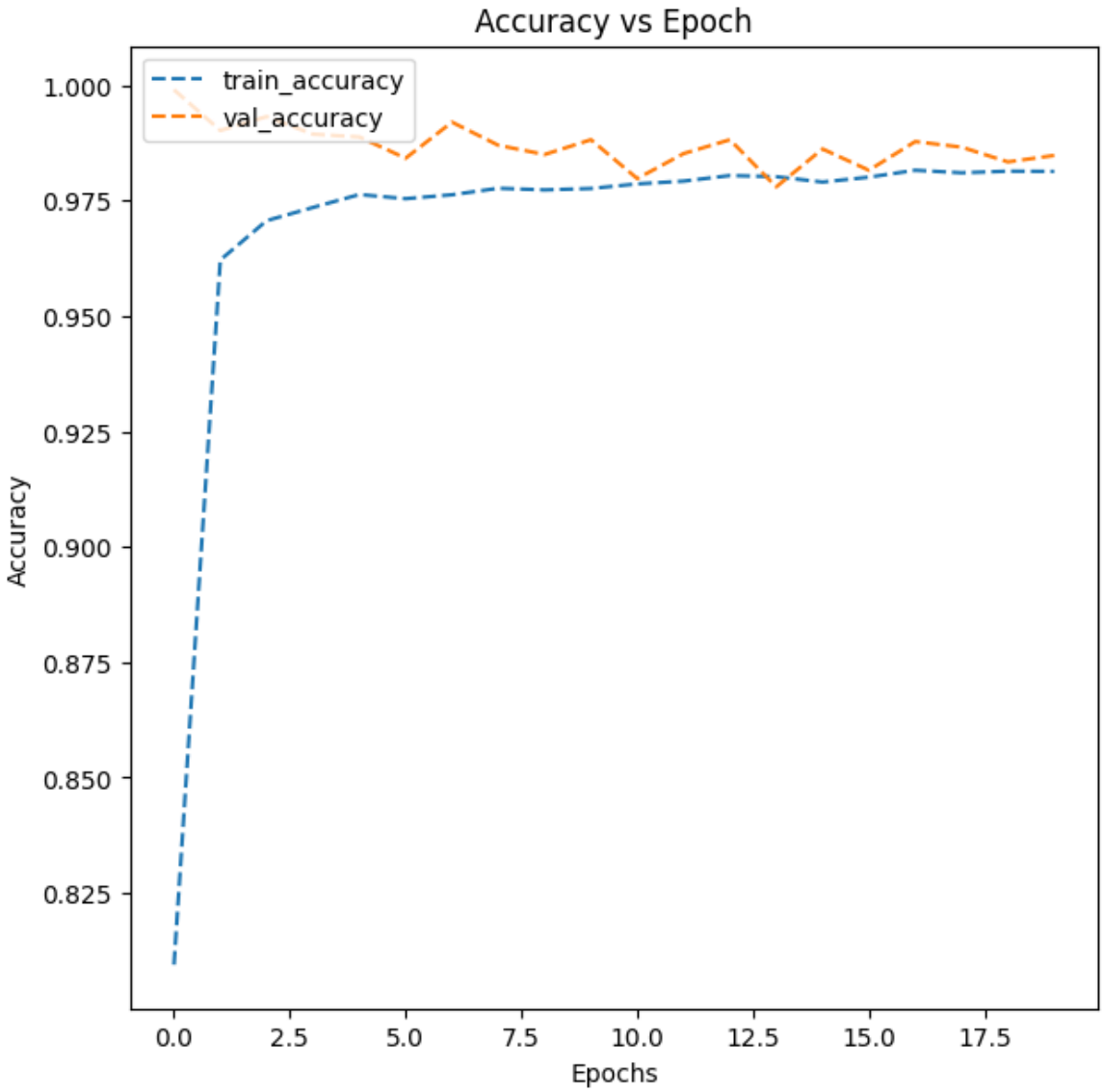
## Models Performance Stats

- Accuracy of models ranged from 98% to 92%.
- False Negatives of the model for raw counts range from 15 to 63.
- In general, the Val accuracy was higher than Train accuracy.
- The training duration for Transfer model VGG16 was larger than other models due to the complexity of VGG16 model.
- The binary classification model did not improve the accuracy but the False Negative raw count increased.

# Model 3 Summary

- 21 layers
- Convolution Layers with LeakyRelu activation – 4 layers
- Maxpooling (improve feature selection) – 4 layers
- BatchNormalization(improve convergence and training) – 4 Layers.
- Dropout Layer(reduce chances of overfitting) – 4 Layers
- Flatten Layer – Convert a CNN output to an ANN input.
- ANN Dense layer – 2
- Final Dropout layer.
- Output Layer with 2 output variable and Softmax activation
- Model compiled with Loss function  of “Categorical Crossentropy”, optimizer of “Adam”.

- Model Accuracy on Test Data:- 0.9846153855323792
- Model 3 – Average Recall and Precision:- 98%
- Model 3 – Classification Matrix values – False Negative Raw Count of 15
- Model 3 – Accuracy Plot



## Choosing the Final Model:

- Based on the testing of the models, Model 3 provides the optimal solution.
- The data augmentation did not improve the model accuracy.
- The use of VGG16 transfer model did not improve the model accuracy and thus would not be recommended.
- The accuracy of VGG16 was low but it could be due to smaller data set and complexity of the model.
- The other factor in choosing a simpler model was the lower False Negative numbers.

## Executive Summary:

- The model chosen has optimal accuracy, has a good performance.
- The False Negatives and False Positives as the minimum.
- The Validation accuracy seems to converge with larger epochs which indicates a good model.
- Improving data in training will address the val accuracy being higher than training.
- The model is not resource intensive and can be implemented easily.