

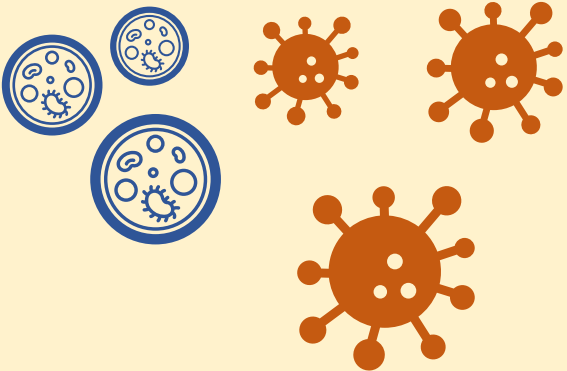


Cell images prediction

Rayees Ahamed

Capstone Project – Three

Springboard Data Science Career Track (2021 -2022)



The Problem


- Infectious diseases such as malaria & dengue are caused by parasites like viruses
- Body cells infected with parasites called – Parasitized cells
- Cells that are not infected called – Uninfected cells
- These cell images can be captured using microscopes

“ Can we predict the infected and uninfected cells from microscopic images?”

Goals of the project

- Labelling the two classes of images
- Extracting the features using image processing
- Training the images on a suitable neural network model
- Predicting the test images using the final model

Data wrangling



Setting the path
for images

Labelling the
images

Converting to
series and
DataFrame

Shuffling the
DataFrame

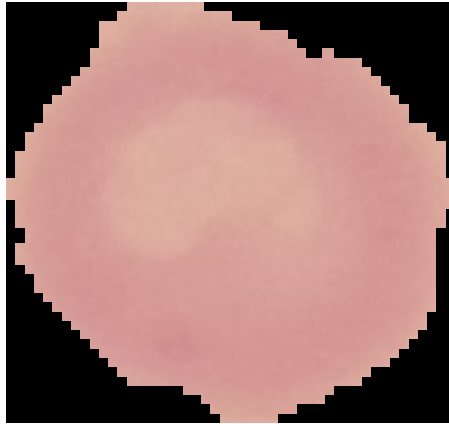


Exploratory Data Analysis

Image visualization



Parasitized cell

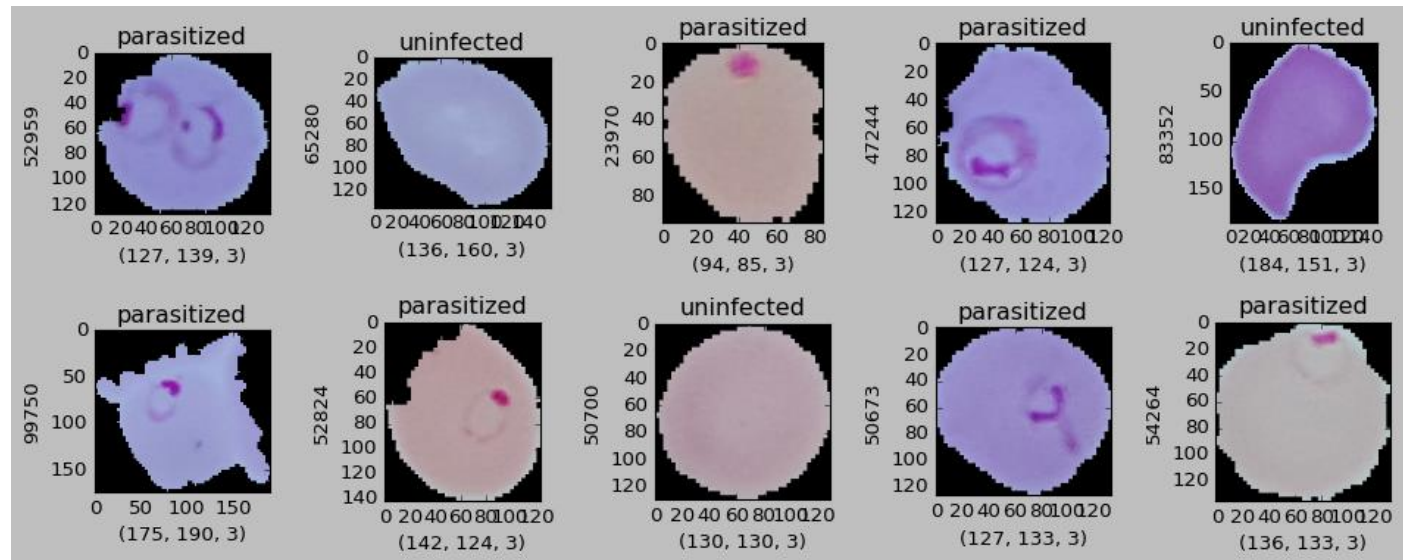


Uninfected cell

- Images were labelled with two classes
- Labelled images verified by matplotlib visualization

Image classes (binary):

- 0 – Parasitized
- 1 - Uninfected



Extracting image features

Image edge detection

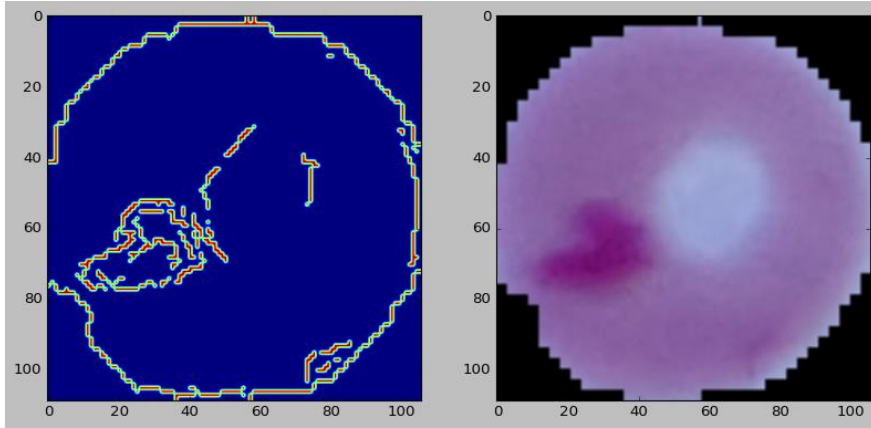
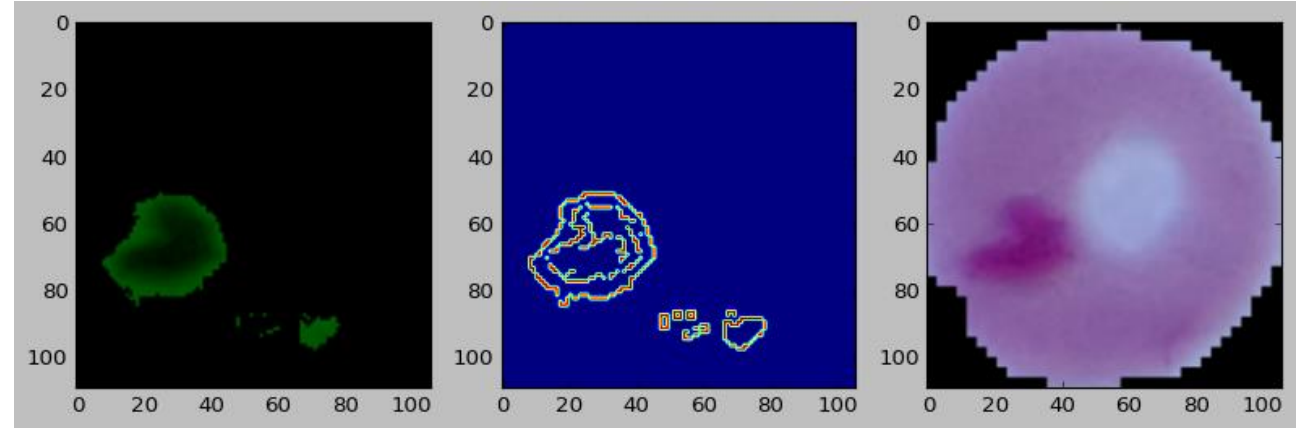
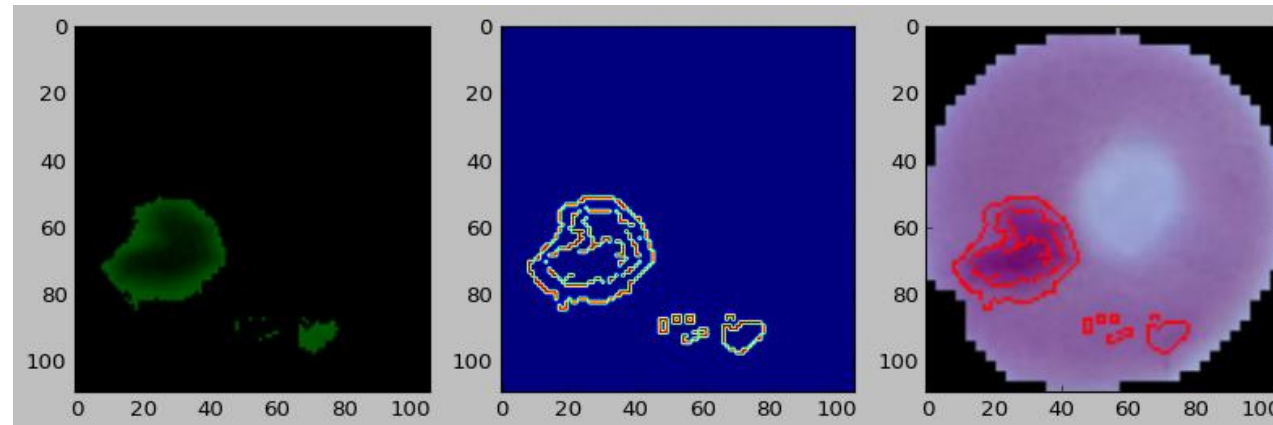


Image thresholding

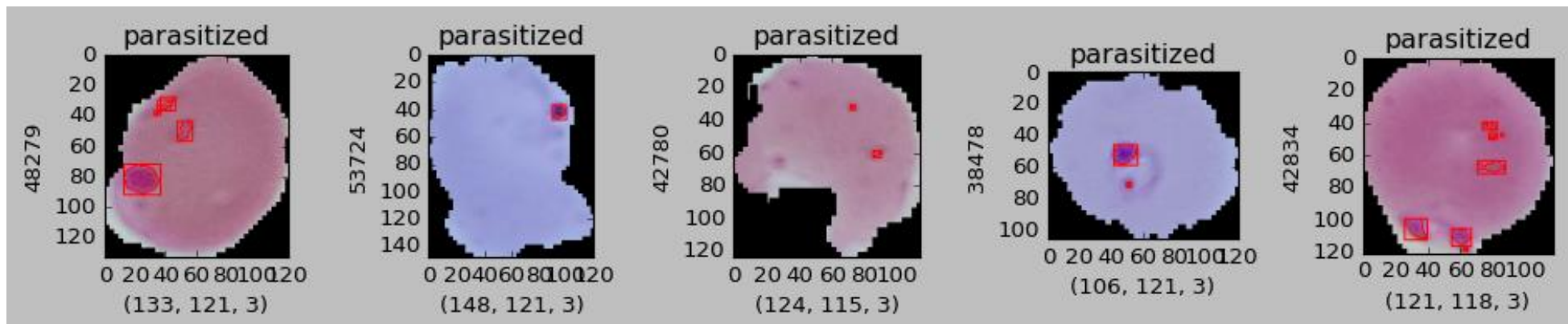


Finding contours

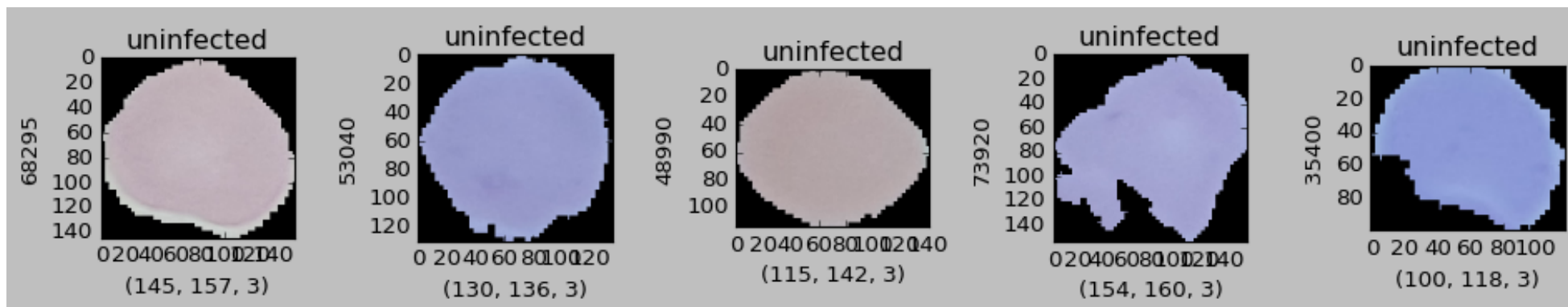


Applying features to the dataset

Parasitized set

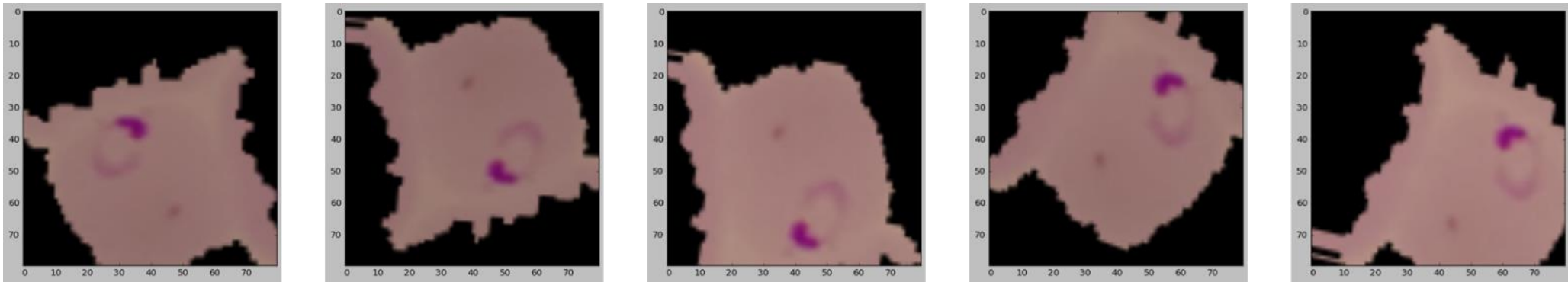


Uninfected set



Preprocessing

Image augmentation



Generator Structure:

- Keras Image Data Generator
- Parameters used:
 - Rotation
 - Zoom
 - Brightness
 - Horizontal / vertical flips
 - Width / Height shift

Modeling - Convolutional Neural Network (CNN)

Model Architecture

Model: "sequential"

| Layer (type) | Output Shape | Param # |
|---|----------------------|---------|
| separable_conv2d (Separable Conv2D) | (None, 254, 254, 12) | 75 |
| batch_normalization (Batch Normalization) | (None, 254, 254, 12) | 48 |
| max_pooling2d (MaxPooling2D) | (None, 127, 127, 12) | 0 |
| separable_conv2d_1 (Separable Conv2D) | (None, 127, 127, 24) | 420 |
| dropout (Dropout) | (None, 127, 127, 24) | 0 |
| max_pooling2d_1 (MaxPooling2D) | (None, 63, 63, 24) | 0 |
| time_distributed (TimeDistributed) | (None, 63, 1512) | 0 |
| bidirectional (Bidirectional) | (None, 63, 64) | 395520 |
| bidirectional_1 (Bidirectional) | (None, 63, 64) | 18816 |
| flatten_1 (Flatten) | (None, 4032) | 0 |
| dense (Dense) | (None, 256) | 1032448 |
| dropout_1 (Dropout) | (None, 256) | 0 |
| dense_1 (Dense) | (None, 1) | 257 |

=====
Total params: 1,447,584
Trainable params: 1,447,560
Non-trainable params: 24

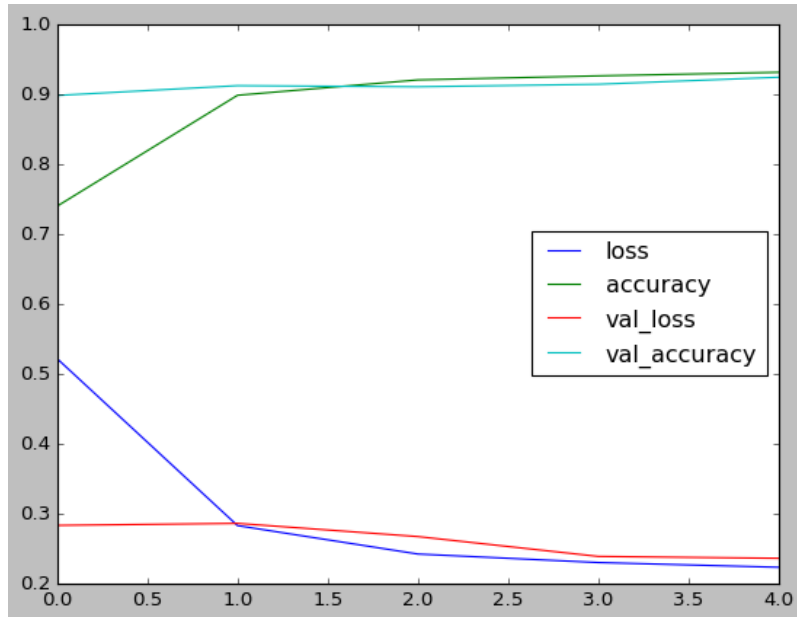
Model
training

Performance
metrics

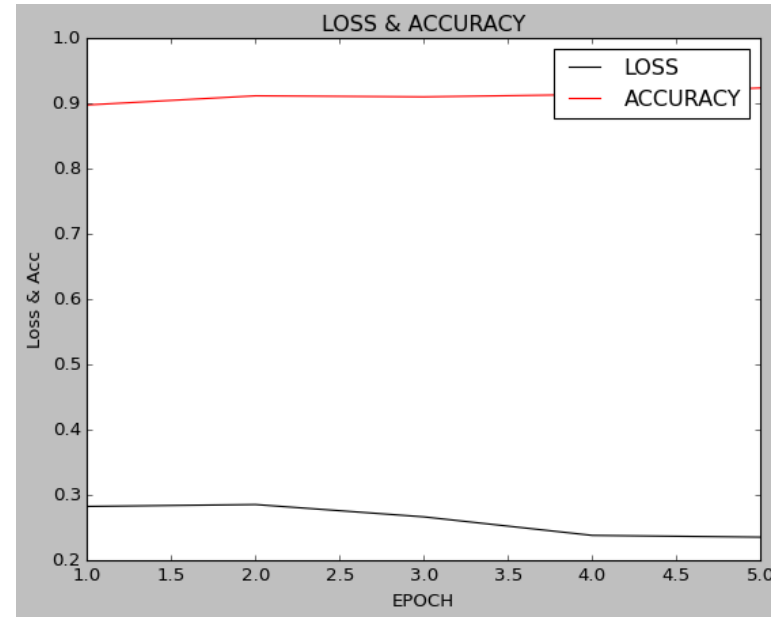
Prediction

Model evaluation

Training vs validation

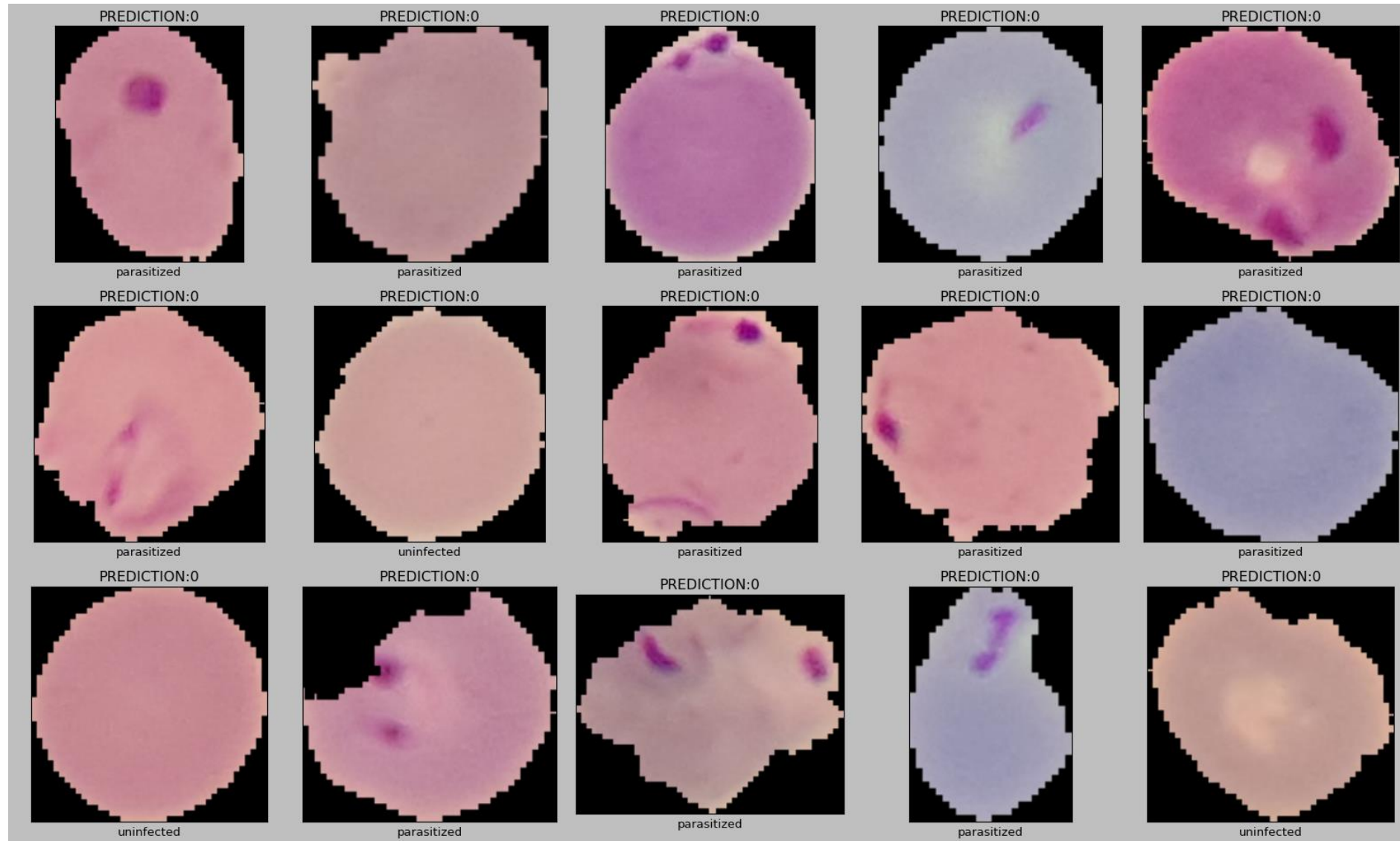


Loss & accuracy vs Epochs



| | Accuracy | Loss |
|----------------------|----------|------|
| Training image set | 0.93 | 0.22 |
| Validation image set | 0.92 | 0.23 |
| Test image set | 0.92 | 0.24 |

Prediction on test images



Summary

- CNN model successfully classified the cell images classes
- High accuracy (94%) and minimal loss (0.2) obtained
- Very few images misclassified into other classes – can be rectified using larger sample size
- Other models can be also tested in future for model comparison study.

Acknowledgements

