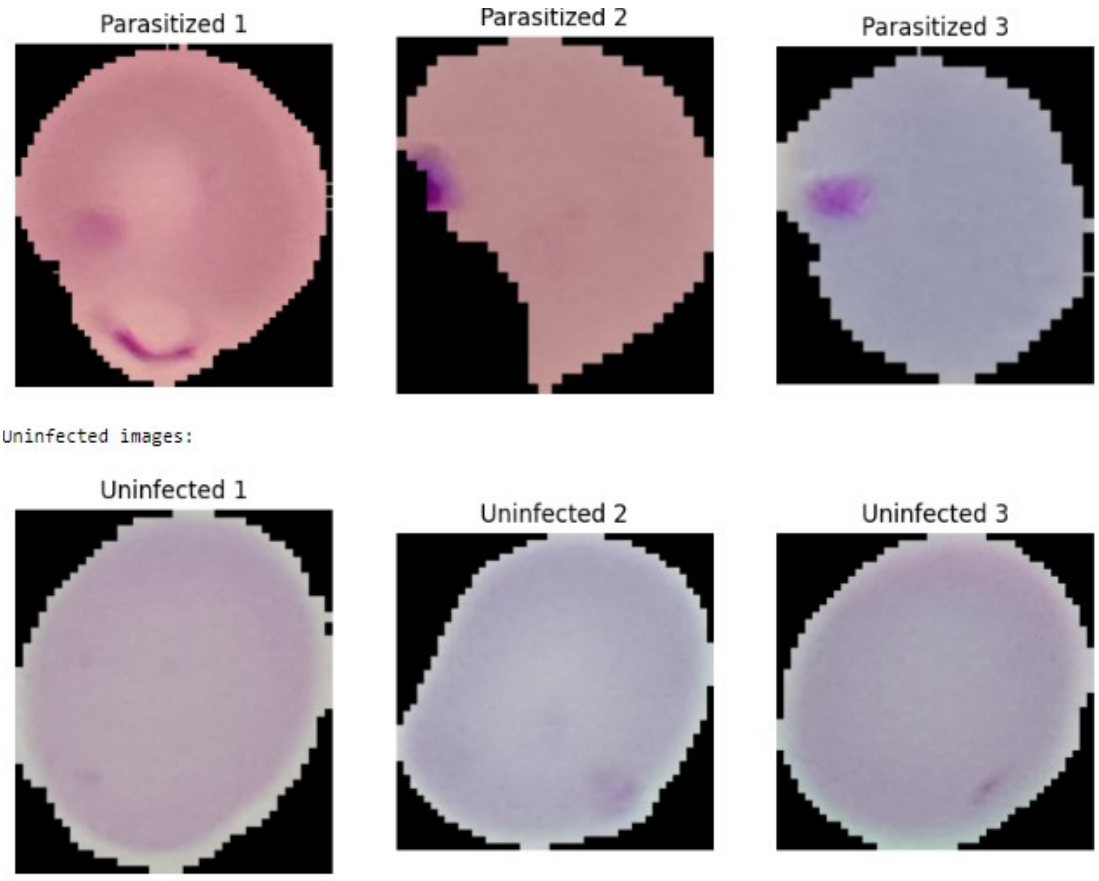
Parasitized vs. Uninfected: A Deep Dive into Image Classification Using Convolutional Neural Networks

Dataset:

- The dataset consisted of images featuring two distinct cell types. 13,780 images were available for each class.
- Parasitized cells had visible darker spots on their cytoplasm, and uninfected cells exhibiting clear cytoplasm.
- Below images show the visible difference between parasitized cells and uninfected cells.



CNN Architecture:

- The architecture consists of three convolutional layers
- First layer would have 32 filters, second layer would have 64 filters and third layer would have 128 filters
- In an attempt to increase the performance of the model further, normalization and dropout layers were added to the architecture.

Model Training:

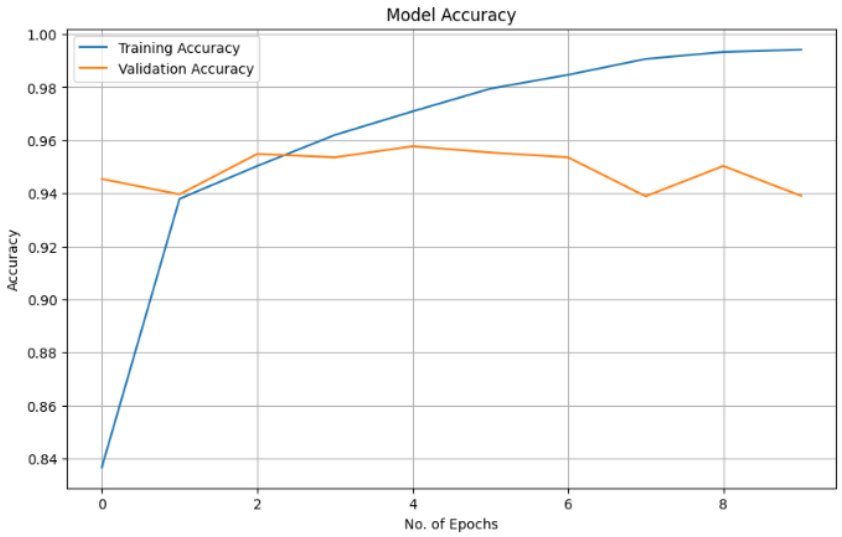
- “adam” optimiser was used during compiling the model
- Model was trained for 10 epochs with batch size of 32.

Parasitized vs. Uninfected: A Deep Dive into Image Classification Using Convolutional Neural Networks

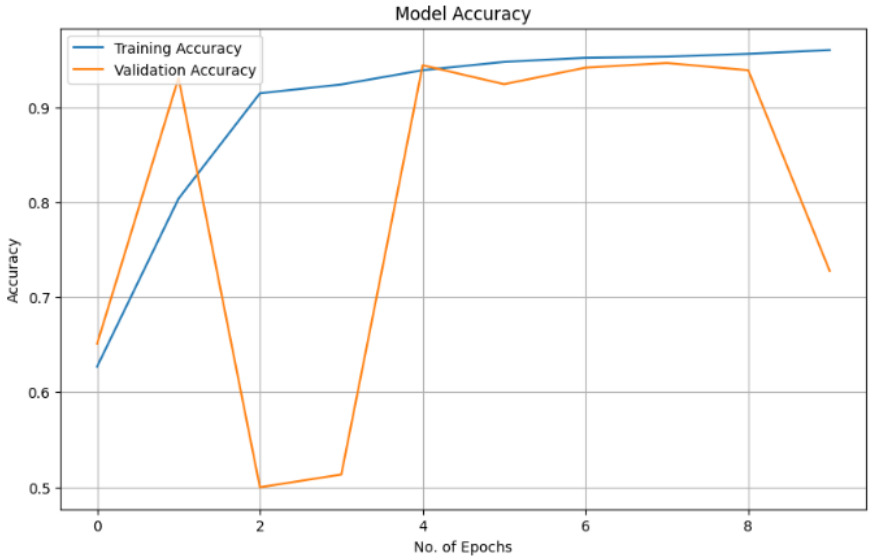
Model Performance:

Model Accuracy:

- Model accuracy for trial 1 is shown in below pot.



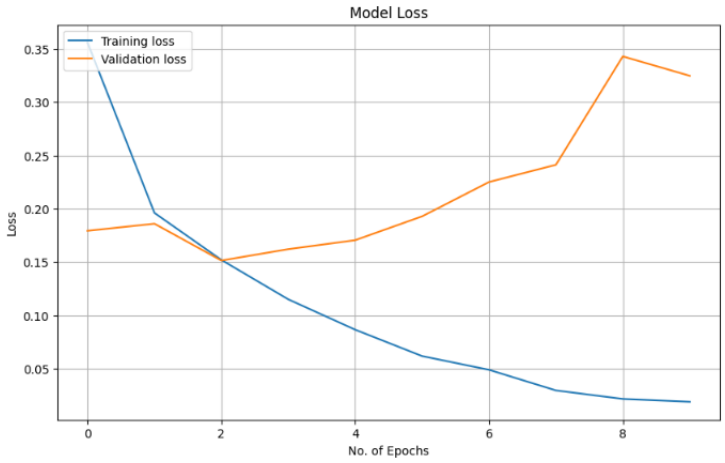
- Model has approximate 99% accuracy on training dataset and nearly 94% accuracy on validation dataset.
- Model accuracy for trial 2 is shown in below pot.



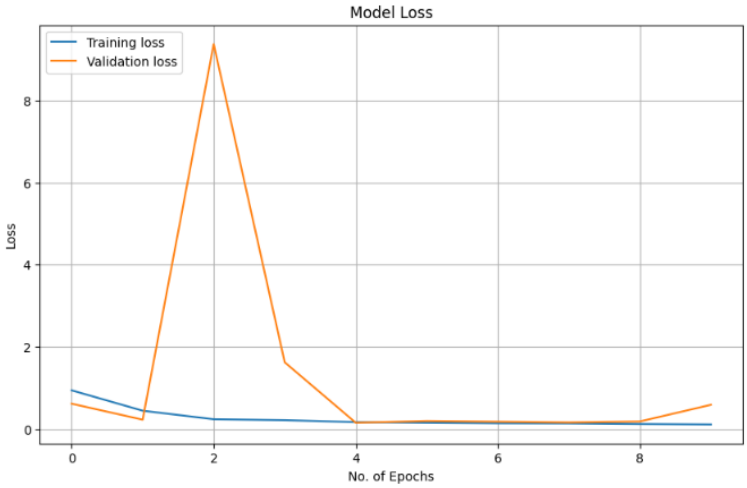
- Model accuracy for trial 2 is below 85% for validation dataset.

Model loss:

- As the number of epochs increase loss in training reduces but loss for validation increases.
- Model loss for trial 1 is shown in below plot



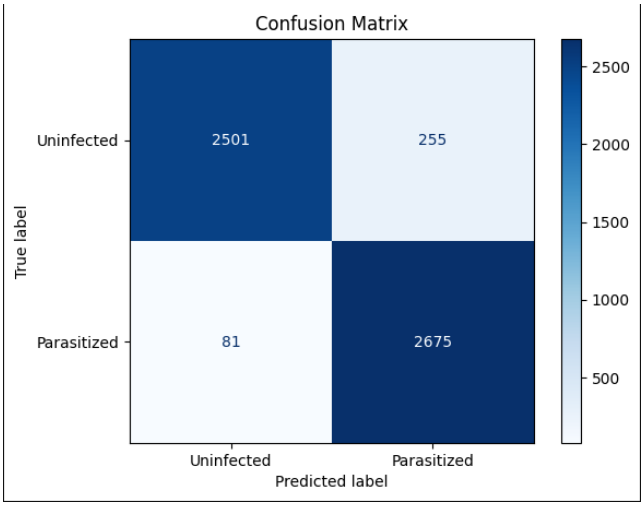
- Model loss for trail 2 is shown in below plot



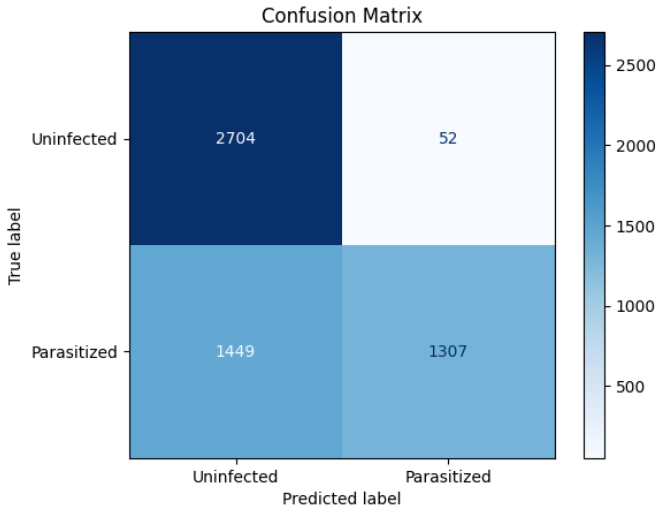
- Model has less loss for training and validation as number of epochs increase but the accuracy on validation dataset is not satisfactory.

Confusion matrix:

- Confusion matrix for trial 1 is shown as follows



- False negatives are lowest among the false predictions, which is required in ML projects in medical domain.
- From confusion matrix it is clear that the model is accurately predicting 94% of the data correctly.
- Confusion matrix for trial 2 is shown as follows



- The negatives have drastically increased, false negatives being the highest.

Conclusion:

- From above plots and confusion matrices we can conclude that batch normalization and dropout is reducing the performance of the model.
- While the model has less loss for training and validation as number of epochs increase the accuracy on validation dataset is not satisfactory.
- Therefore, the fist model (Trial 1) has been used for deployment.

Scope for Improvement:

- Due to time constraints on GPU usageon Google Colab only two trials were performed.
- More trials can be performed by using L1and L2 regularisation techniques.
- Trails can also be performed by varying batch sizes and number of epochs.