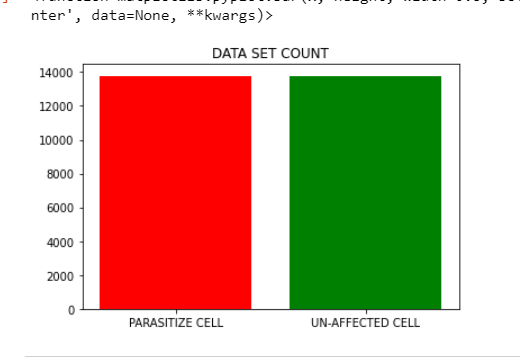
**Malaria Cell Classification Report**

1. **Import and Exploration of dataset**:

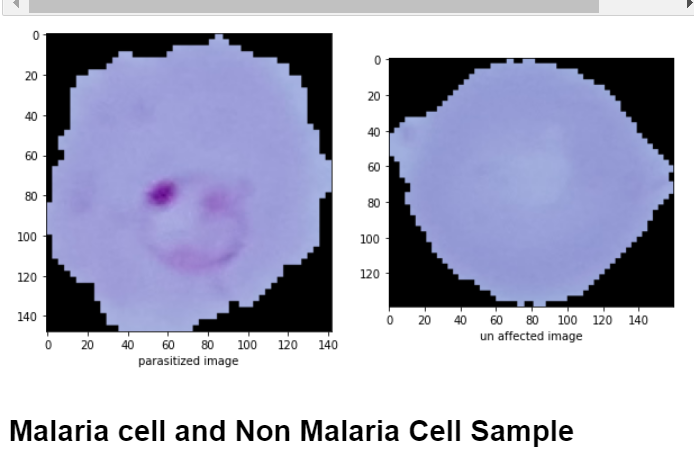
The image data are download in two separate folders (Parasitized folder and Uninfected folder). With the help of Pathlib module in python, accessing of widow path is easy and importing of those images into the Jupyter Notebook with the help of ‘cv2’ module in python. Images samples is stored in a list variable ‘X’ and the corresponding label his store in a list variable ‘y’. Based on this work the Parasitized cells are label ‘0’ while the unaffected cell is label ‘1’, its essential to convert the ‘X’ and ‘y’ list into a ‘Numpy’ array module for easy computation and manipulation of matrix (image matrix data).

Furthermore, A total image of 27,558 images dataset are loaded for both the parasitized and uninfected cell image, 13,779 images belong to the Parasitized image class while the uninfected class contain 13,779 images as well. This shows that the dataset is balanced for both class. The figure 1 below shows a bar chart representation of the number of image samples from both class.



**Fig 1. Sample Cells Image size from both class**

Considering the fig 2. Show a sample from each class by using the MathPlotLib.pyplot module in python. Hence, it was identified that the images are not of the same dimension. Further exploration is carry out by checking the first five image shape (size and dimension) of each class, based on the result its strongly confirmed that all the images are of different shapes and dimensions.



**Fig 2. Cell Images from both class (unequal dimension)**

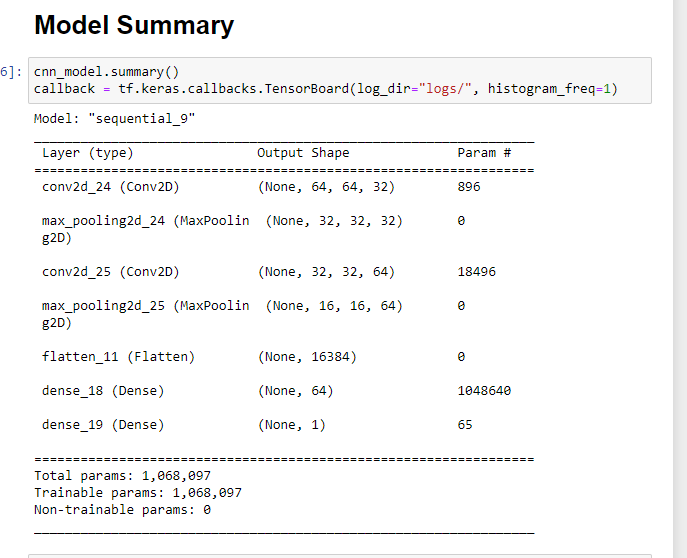
1. **Malaria Cell Data Preprocessing**

Based on the un-uniformity of image size and dimension its essential to reshape or resize all the image shape. With the help of ‘*resize’* function in ‘cv2’ module the images where all resize to a size of (64,64). In addition, image is required to be scale (scale 0-255 to 0-1) or normalize in other to improve machine or deep learning accuracy while training and making predictions. Finally, the images dataset as to be divided into training and testing, based on this work the training images and label samples is 20,668 (75% percent of the entire data samples) with image size and dimension of (64,64,3), while the testing image sample and label is 6,890 (the remaining 25% percent of the entire data samples) with the same image size and dimension of (64,64,3).

1. **Building the Convolutional Neural Network (CNN MODEL).**

Convolutional neural network belongs to one of the deep leaning family. It’s mainly used specifically for image processing, detection and classification, has a result of this the CNN is selected for building the Malaria Cell image classifying problem. The deep neural network was built with two convolution layer and MaxPooling2D layer, the convolution layer is parameterized with filter size of 32 (No of feature extraction), Kernel Size or feature cell of (3,3), input shape of (64,64,3) images and activation function used in those layers is ‘ReLu’, the MaxPooling2D is used for size reduction in other to improve learning accuracy. Furthermore, after a series of convolution (feature extraction) and MaxPooling2D (dimensionality reduction) a flatten layer is define to convert the 3-dimension image into 1-dimension image, then this is follow by a fully connected Dense layer of 64 neurons and a output layer of 1 neurons with activation function defines as ‘sigmoid’.

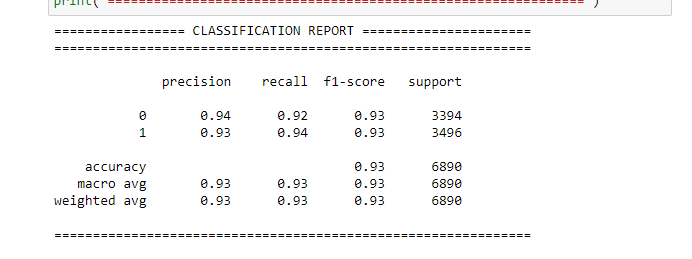
However, the compile by defining the hyper parameters such as loss, optimizer and metrics. Based on the work the output is a binary output (0 or 1) so ‘*binary cross entropy’* is use as the lost function and ‘*adam’* as the optimizer, and the metrics used is ‘*accuracy’.* finally, the model is now compiled. The figure 3 below shows the Model configuration.



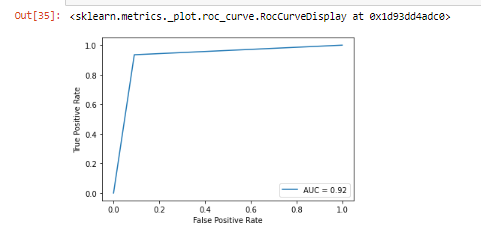
**Fig 3. Model Summary (Configuration of model layers) ….**

1. **Model Evaluation**

Models are evaluated using the popular known metrics in the field of machine learning or deep learning, this include Accuracy, precision, recalls, and F1-scores others include ROC Curve and PR-curve. The classification report and the confusion show the accuracy, precision and recall of the developed model. based on the figure 4 below and accuracy of 93% was achieve alongside with 93.5% precision, 93% recall and a F1-score of 93. The ROC curve shows a good performance across all threshold and the PR curve show a good precision and recall value of ‘1’.

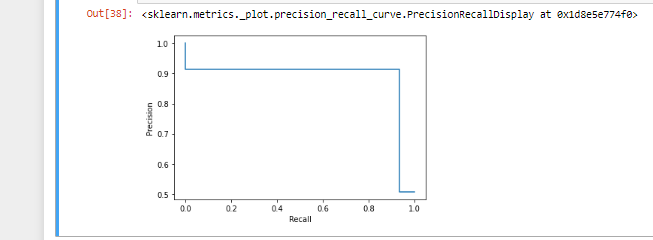


**Fig 4. Classification Report**



**F**

**Fig 5. The ROC curves**

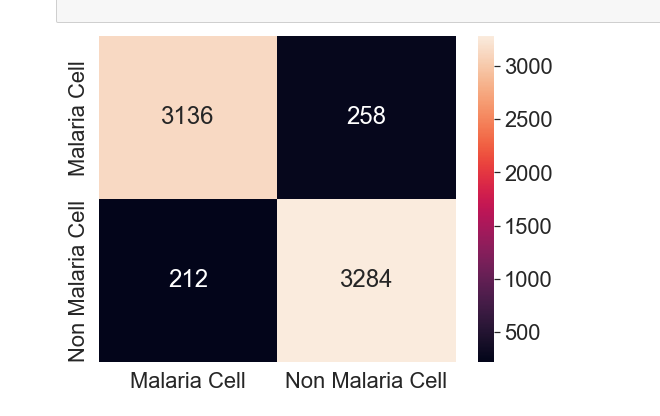


**Fig 6. PR curve**

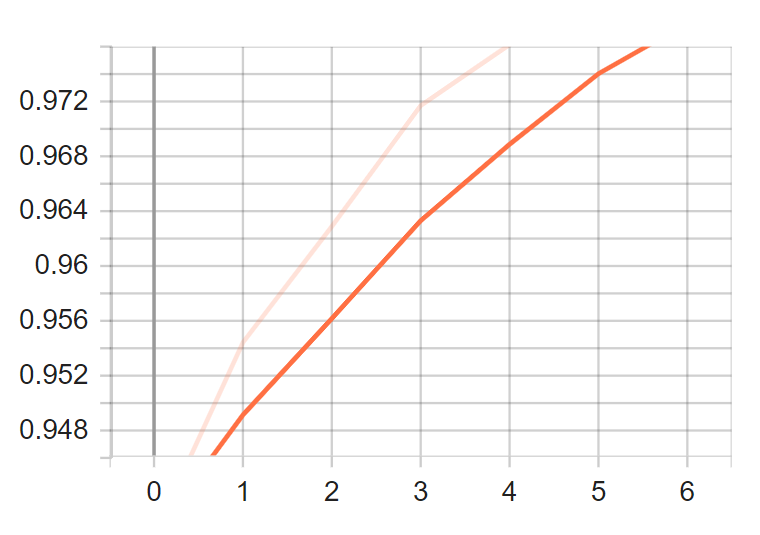
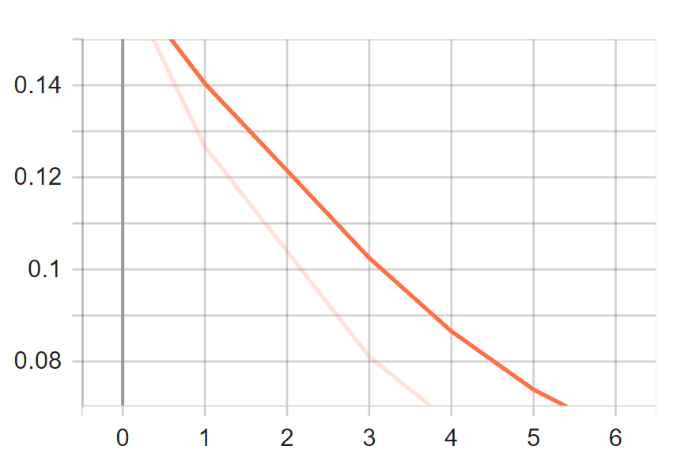
1. **Fin-tuning Model**

The parameter of the CNN model is adjusted in other to get a better and more accurate result. This include adjusting of the Filter (determine the number of features to be extracted) parameter of the Covn2D layer, number of neuron to drop, loss hyperparameter, number of epochs and the likes. Finally, with 32 filter, 7 epochs, binary cross entropy as loss, 0 dropout and the Convolution layers, the model perform best.

1. **Model Evaluation**



**Fig 7. Confusion Matrix**

Based on the figures 7 above, thus the confusion matrix is can be deduced that the 3,136 Malaria images are correctly classify while 258 Malaria images are wrongly classified. However, 3,284 Non-Malaria cell images are correctly classified and 212 Non-Malaria cell are wrongly classified.

**Accuracy**

**Loss**

**Epochs (Iteration)**

**Epochs (Iteration)**

**Fig 8 Accuracy level with epochs fig 9 Loss with epochs**

In consideration of fig 6. Which shows a line graph of accuracy at every epoch or iteration during the training process, based on the graph its identified that the accuracy level start with approximately 94% accuracy and for every epoch there is an increase in accuracy level. It’s also identify from the graph that the training takes 7 epochs or iteration. While figure 7 show that at every epoch there is a significant reduction in loss. This shows that for every iteration the deep learning algorithm is able to learn new features.

It is highly recommended that employing feature extraction techniques such as Histogram of Oriented Gradient (HOG), Binary Local Pattern (LBP) or Binary Robust Independent Elementary Feature (BRIEF) and the like can drastically improve the accuracy of the prediction model.