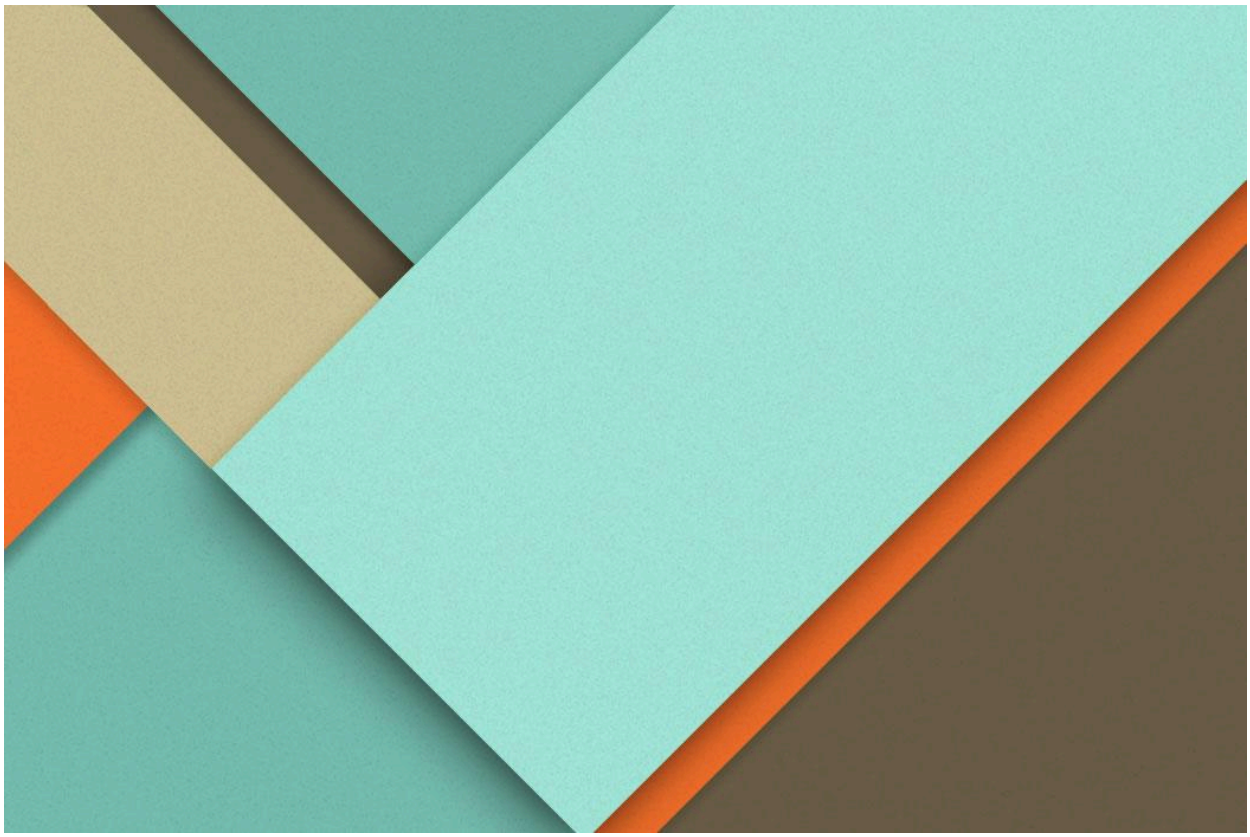


MALARIA DETECTION WITH MACHINE LEARNING

Martha Sharon Murugi Maina

15th May, 2024

Overview

The following is an overview of the malaria detection assignment with machine Learning

Tools & Software

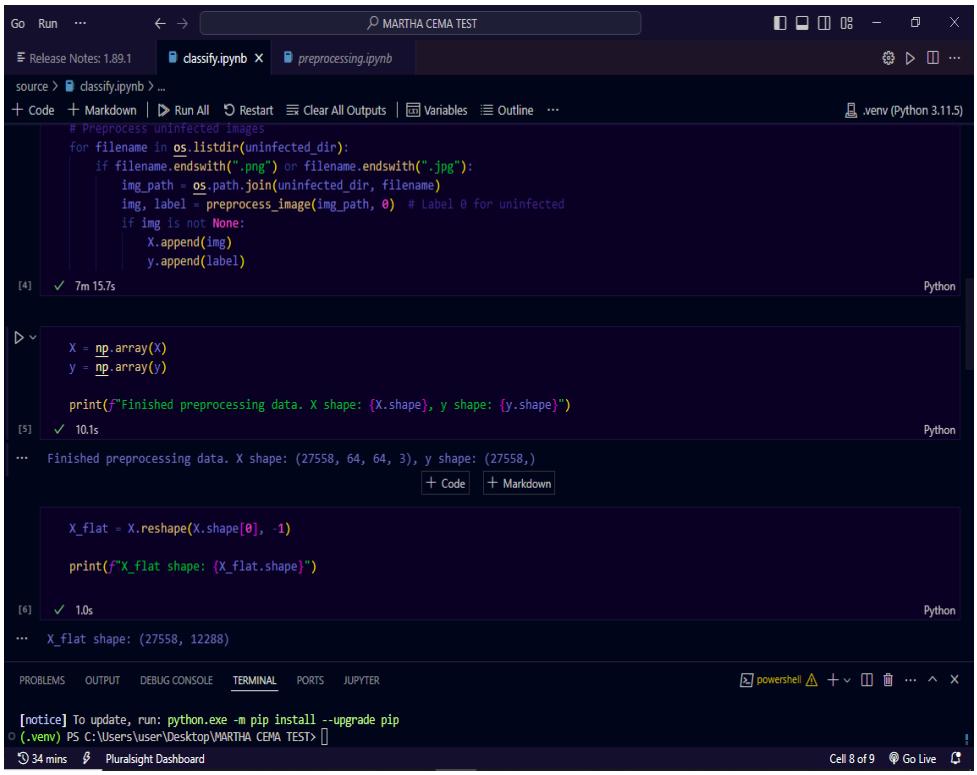
1. Visual Studio Code
2. Anaconda
3. Google Sheets
4. Google Docs
5. Python
6. Scikit-Learn
7. Pandas
8. Numpy
9. Pillow

Data Collection & Analysis

The dataset was downloaded from Tensorflow Datasets. I analyzed it using Google Sheets. There were no duplicates or white spaces. However, there were two empty cells which I cleared.

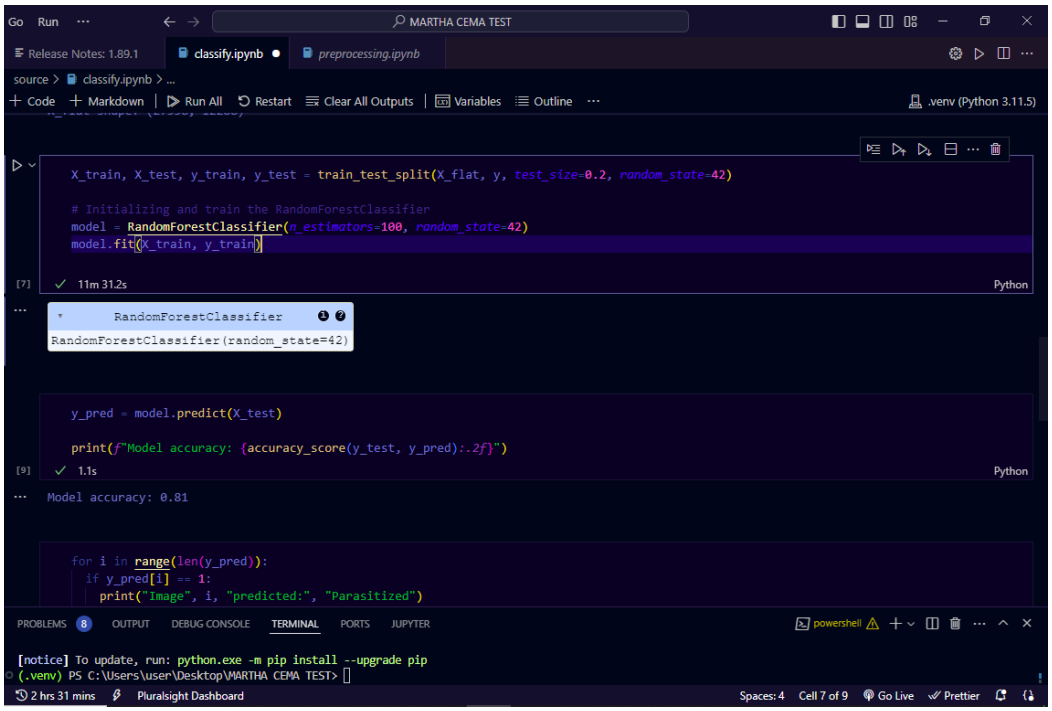
Data Pre-processing

I used Numpy and Pillow to read, resize and normalize the images. The images were in different subfolders and so preprocessing was for each individual folder.



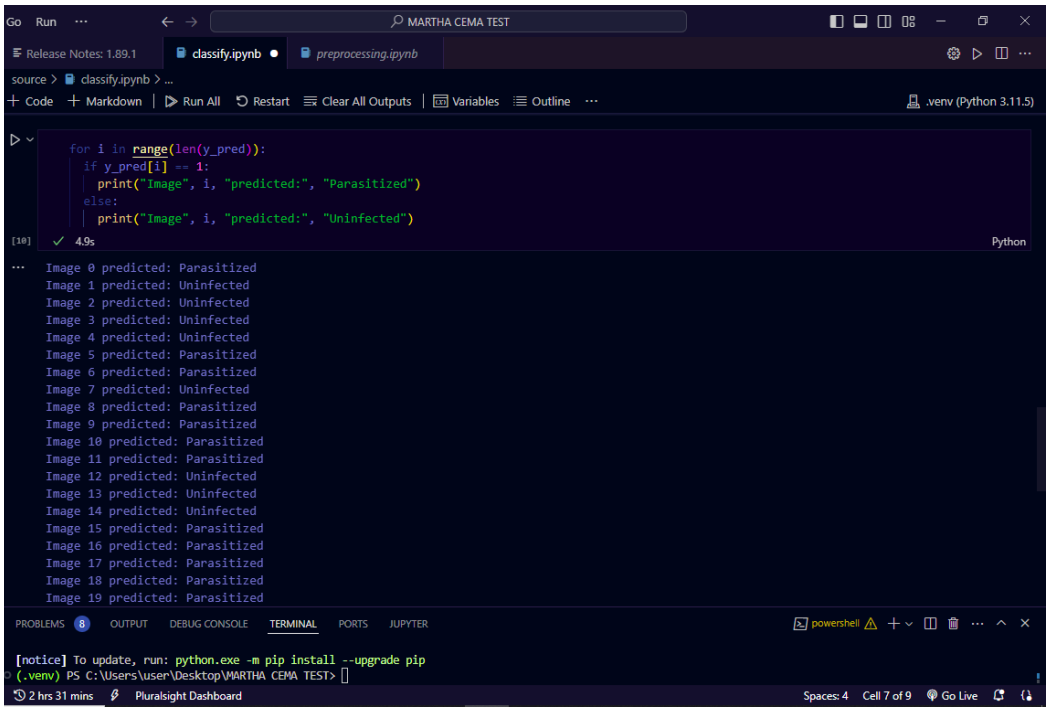
Classification Model & Training

I. RandomForest Classifier



I used RandomForest Classifier for its ease of use and it served as a strong baseline model. Upon training, the model achieved 0.81 model accuracy.

II. Prediction Outcomes



III. Suggestions and Recommendations

To reveal specific types of errors a Confusion Matrix would do best as it shows how many instances from each class were predicted correctly (diagonal) and incorrectly classified as other classes (off-diagonal elements).

Conclusion

This study investigated the application of a machine learning model for malaria parasite detection in blood images. A Random Forest classifier model was trained on a dataset of preprocessed blood cell images labeled as parasitized or uninfected. The trained model achieved an accuracy of 81% on a separate testing set.