

Case Study: Detecting Malaria

Topic: Identifying cells that have been infected with Malaria

Who is responsible for completing this case study: UVA 2nd year students-you guys are the future! Take what you have learned from your data science classes and show your class just how much you have learned.

Context: Nearly half of the world is at risk for malaria transmission. As of 2021, there were 247 million malaria cases and 619,000 of those cases led to deaths. Malaria is diagnosed by examining the patient's blood under a microscope after it has been stained (to make the parasite more distinct). The picture on the left shows what an infected cell looks like and the picture on the right shows what an uninfected cell looks like.

Infected cell:



Uninfected cell:



Motivation: Why take on this project? In order to treat malaria, it must be recognized as quickly as possible so that it can be treated quickly and more lives can be saved.

Prompt: You have been hired by the CDC to make a model that can look at images of cells and predict whether they have been infected with malaria or not. They have provided the dataset (originally from National Institutes of Health) for you to train your model on and some hints to get you started. However, other than the dataset, you have full control of this project-you make the calls. It is your responsibility to determine what the best modelling approach is and come up with the best model possible. Be ready to research-the CDC is counting on you (but also have fun).

Deliverable: Produce a model that can detect cells infected with malaria and report its accuracy and f1 score.

Hints: The dataset you will be using is an image dataset-look into convolutional neural networks and how they work. The tricky part is getting your folders set up appropriately to pass in the data to your model, you shouldn't have to do too much data cleaning for this particular dataset. Have fun with this! Even after you have a working model, keep playing with the hyperparameters to see if you can get better accuracy rates.