

Predicting Malaria

Malaria detection through blood cell images analysis using deep learning

Predicting Malaria

Agenda

- ❖ Motivation
- ❖ The Data
- ❖ Data Preprocessing
- ❖ Modeling
- ❖ Conclusion

Predicting Malaria

Motivation

- ❖ Malaria cases must be recognized promptly in order to treat the patient in time and to prevent further spread of infection in the neighborhood via local mosquitoes. Malaria is considered a potential medical emergency and should be treated accordingly.
- ❖ Delay in diagnosis and treatment is a leading cause of death in malaria patients. Approximately 400, 000 deaths every year world wide which is mostly affecting poor areas.

Predicting Malaria

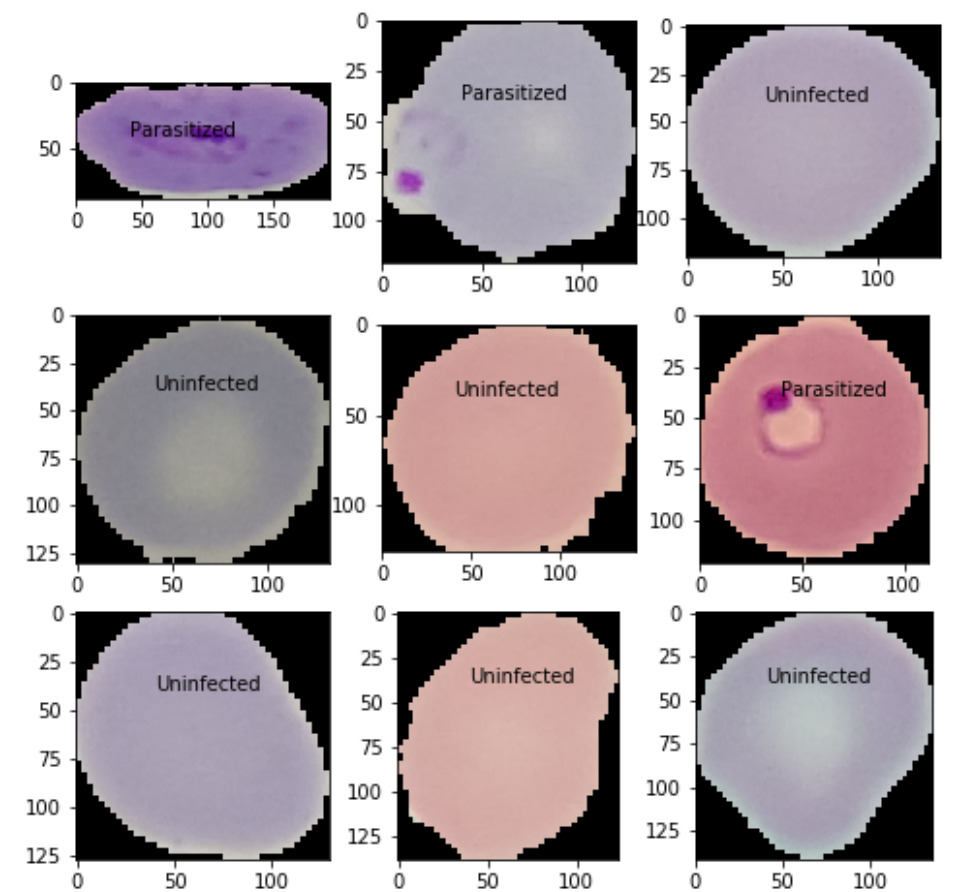
Motivation

- ❖ The way malaria is diagnosed is a bit tricky as it requires both clinical diagnosis and microscopic diagnosis. Clinical diagnosis phase involves checking patient's symptoms such as fever, chills, sweats, muscle pains, etc.
- ❖ Microscopic diagnosis is examining patient's blood sample under the microscope to identify parasites.
- ❖ Building up a model to help diagnosing malaria cases can help automating the process and clinicians so that they can spend more time on treating patients.

Predicting Malaria

The Data

- ❖ The dataset used in this project is obtained from the U.S. National Library of Medicine, which is set of blood smear images collected by researchers.
- ❖ The images were manually annotated by experts at the Mahidol-Oxford Tropical Medicine Research Unit in Bangkok, Thailand.
- ❖ The dataset contains a total of 27,558 cell images with almost equal instances of parasitized and uninfected cells.



Raw cell images data

Predicting Malaria

Data Pre-Processing

- ❖ In order to feed image data to machine learning models we use a few preprocessing methods such as scaling, resizing, transforming, rotating and shifting. The figure below show raw image samples before preprocessing.
- ❖ Dataset is downloaded from NIH website in CSV format which includes labeled cell images. We first write a source code to divide dataset into three sets those are training, validation and test sets.
- ❖ Training set contains 80 % of the whole data while test and validation sets are sharing 10 % respectively. After splitting dataset, we seek for the best size to resize images so that our algorithm would work at optimum pace

❖

Predicting Malaria

Data Pre-Processing

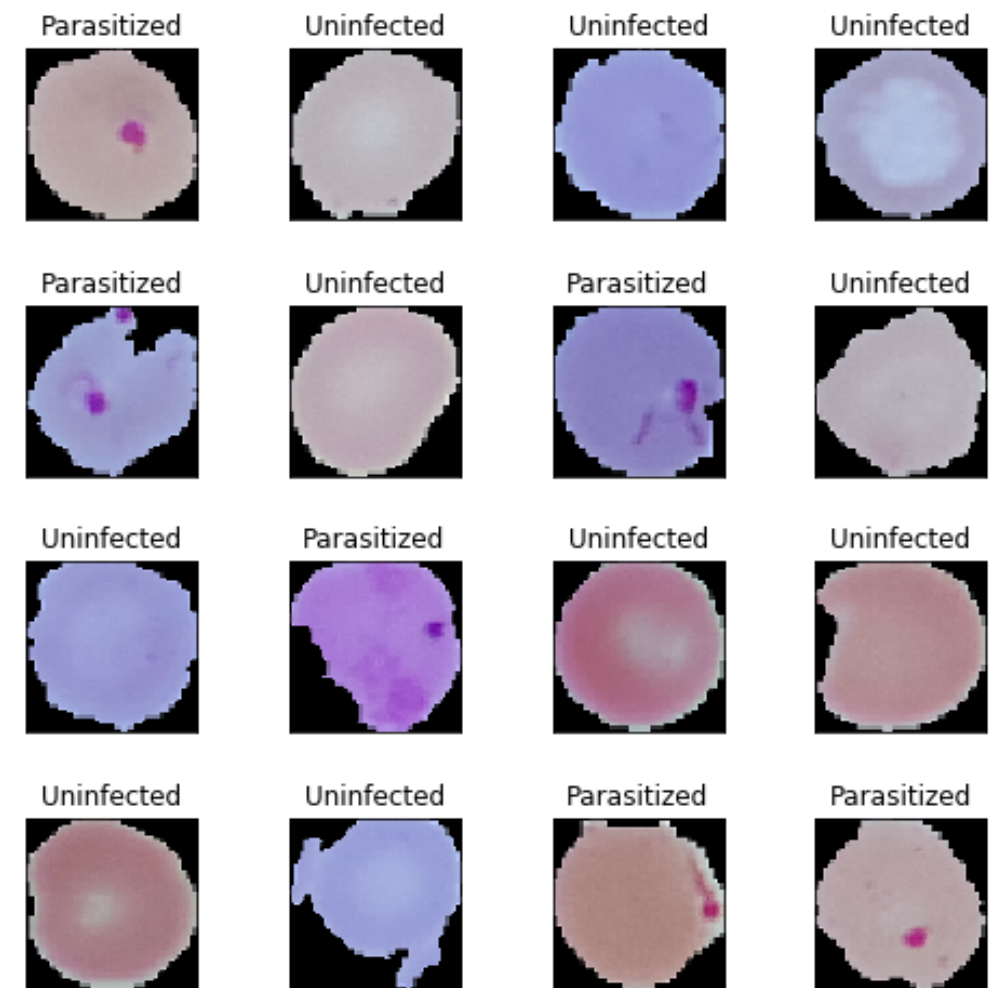
- ❖ Average pixel dimension was obtained as [133.16, 132.61, 3], while median pixel dimension was [130, 130, 3].
- ❖ However, since we ran our models on a CPU, we picked even smaller dimensions as such [64, 64, 3] to save time on modeling phase.

	SampleSize	Height	Width	ColorChannel
TrainData	22235	64	64	3
Validation	2470	64	64	3
Test	6177	64	64	3

Predicting Malaria

Data Pre-Processing

- ❖ The figure illustrates the blood smear images after resizing process. We will also apply transformation, rotation and shifting techniques later in modeling section using Keras data augmentation methods.



Predicting Malaria

Modeling

- ❖ What we expect from machine learning algorithm is to detect blood cell images with parasitizes which can be seen as red dot by a microscope after sample is stained with a contrasting agent to help highlight malaria parasites in the blood cell.
- ❖ If that is manually done by a clinician, it would take a lot of time to count all.
- ❖ There is actually a web application developed by Dr. Carlos Atico Ariza which uses deep learning behind the scenes that screens and diagnose Malaria.

Predicting Malaria

Modeling

- ❖ We picked our baseline model as K Nearest Neighbors (KNN) model to start off modeling.
- ❖ Then Simple Neural Network with two hidden dense layers tested out.
- ❖ Next, Convolution Neural Network built from scratch with getting help of my mentor.
- ❖ Finally, the ResNet model is tested out after a few manipulation from its original version on Adrian's blogpost.

Predicting Malaria

Modeling

KNN

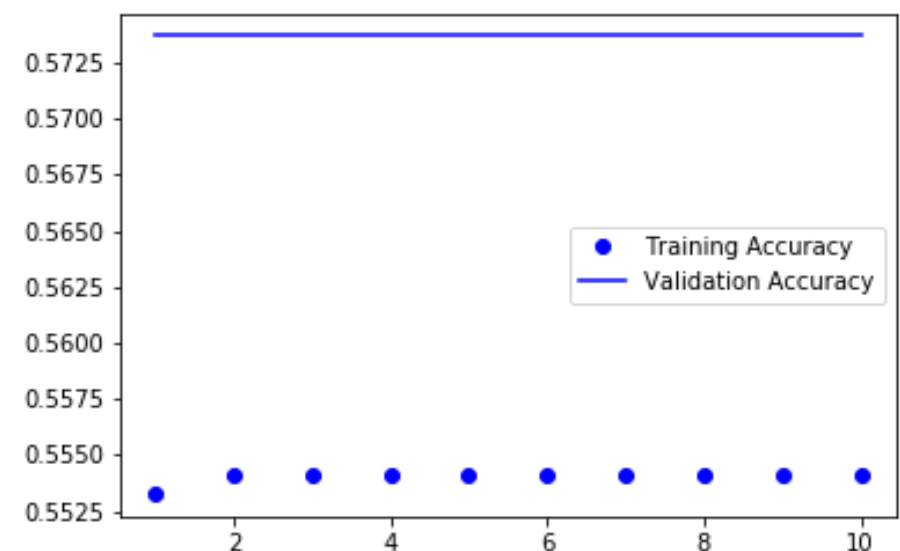
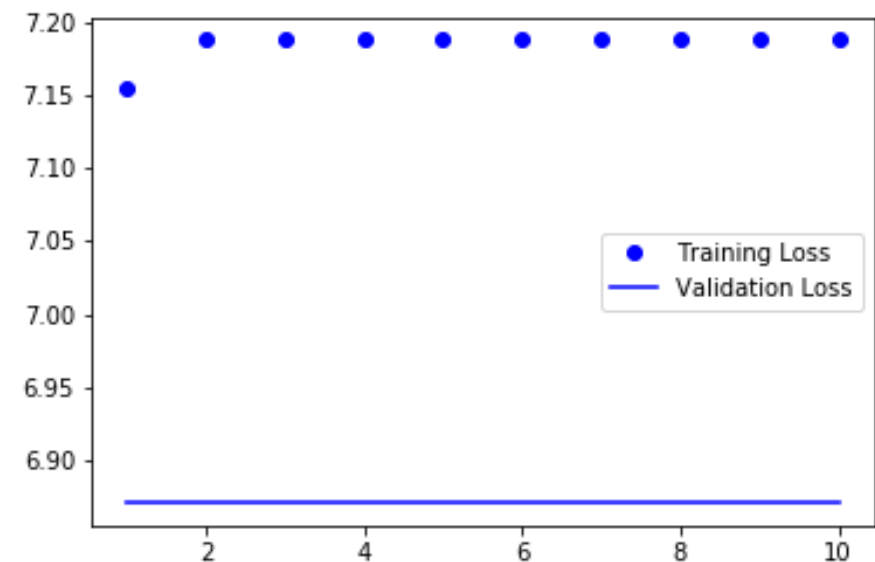
- ❖ We have got 59 % accuracy score with KNN which is not too bad even with this not learning method, given that the probability of randomly guessing the correct class is 50 %.
- ❖ KNN model correctly classified parasitized blood cells 51 % of the time.
- ❖ Looks like KNN did a better job on classifying `uninfected` class based on the Precision score. Apparently, some `uninfected` cells classified as `infected` (False Positive) as we can see from Precision score of `infected` class.

Predicting Malaria

Modeling

Simple Neural Network

- ❖ The model used here consists of a sequence of two hidden densely connected layers having 768 and 384 nodes in each respectively to classify our images. After 5 iteration, trained model has reached about 56 % accuracy which is not really useful.



Predicting Malaria

Modeling

Conv Neural Network

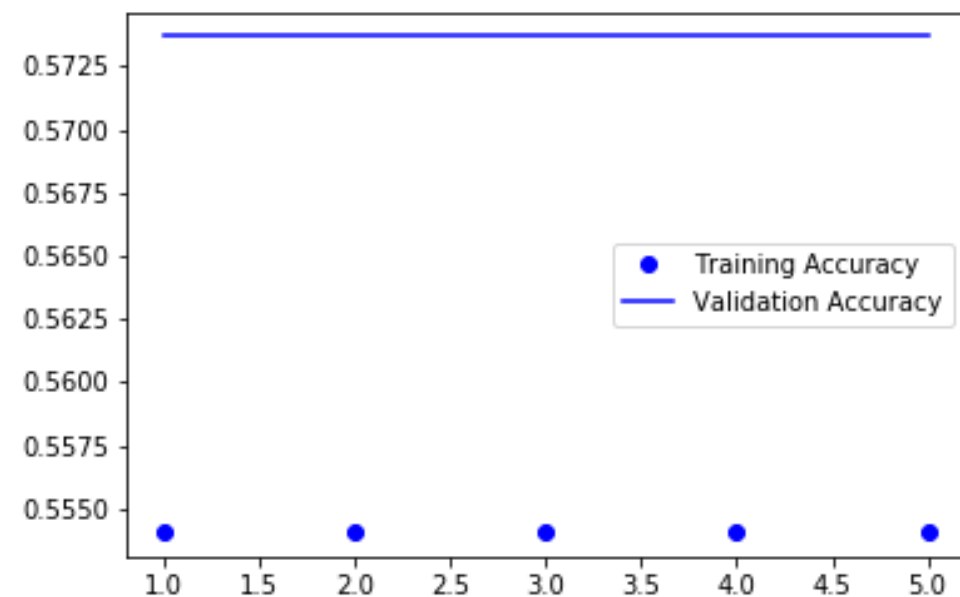
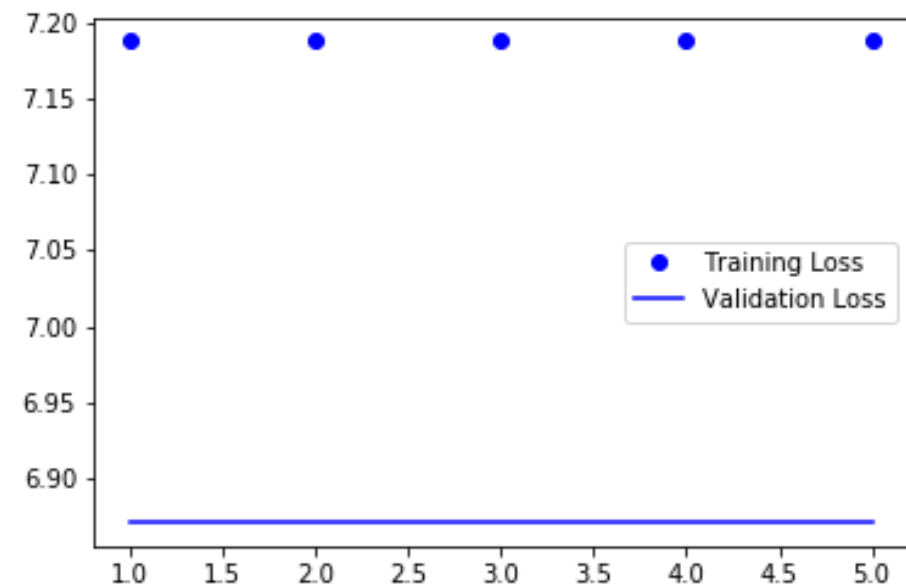
- ❖ Conv2D are convolution layers that will deal with our input images, which are seen as 2-dimensional matrices.
- ❖ In our CNN model kernel size is defined as (3, 3) matrix which is a filter matrix used for blurring, sharpening, embossing, edge detection and more by visiting every pixels in image.
- ❖ We also use dropout to regularize and reduce unnecessary feature dependencies in our network. It helps reducing over-fitting and generalization of the model.

Predicting Malaria

Modeling

Conv Neural Network

- ❖ Max pooling is a down-sampling technique used to reduce the amount of parameters and computation in network and to control over-fitting.
- ❖ We used most common size of (2, 2) pooling layer here in the model.



Predicting Malaria

Modeling

Residual Neural Network (ResNet)

Here is some details about the data preprocessing before feeding data into ResNet model:

- ❖ Batch Size 128,
- ❖ 5 Epochs which took 2 hours to complete with CPU,
- ❖ 0.1 learning rate,
- ❖ (64, 64, 3) image size and
- ❖ also data augmentation technique of Keras is implemented for each dataset.

Predicting Malaria

Modeling

Residual Neural Network (ResNet)

- ❖ ResNet will then perform (3, 4, 6) stacking with (64, 128, 256, 512) CONV layers where the first 64 filters will be applied before reducing spatial dimension.
- ❖ Next, 3 sets of residual modules that will be learning 32, 32 and 128 CONV filters will be stacked.
- ❖ Followed by reducing the spatial dimensions and stack 4 sets of modules where each CONV layers will learn 64, 64 and 256 filters.
- ❖ Finally, we stack 6 sets of residual modules where again each CONV layers learns 128, 128 and 512 filters.

Predicting Malaria

Modeling

Residual Neural Network (ResNet)

- ❖ There is substantial improvement on results compared to KNN, which was our baseline model.
- ❖ Infected class was caught with 98% precision and 94% recall while the geometric mean of them f1 score is 96%. So the ResNet model is almost perfect with a little FP and FN values.
- ❖ High precision scores indicates that there is almost no blood cells that were healthy and classified as infected.
- ❖ Training accuracy of model as mentioned is about 96% and validation accuracy is about 95%.

Predicting Malaria

Modeling

Here is the overall comparison table of model accuracies below.

Model Name	Validation Accuracy
KNN	59%
Simple NN	58%
CNN	58%
ResNet	96%

Predicting Malaria

Thank You
