Capstone Project

Malaria Detection

Milestone 2

Joon Jung

Refined Insights:

1. What are the most meaningful insights from the data relevant to the problem?
   1. Good interpretation of the output from the data
   2. Potential reason for that output
   3. What it means for the problem/business?

Table

Description automatically generatedIn this milestone 2, different deep learning models were tested and validated for its accuracy on predicting the classification of parasitized cell or uninfected cell from its images. First the base model was created to evaluate the performance of a CNN architecture. Later models were modified and validated to build better-performing CNN models. The parameters used to evaluate how good of performance the model can carry were its accuracy, validation accuracy, and f1 score. We need to pay attention to why the f1 score is very important parameter that we use particularly for this project.

In case of this project, false negative would mean we are predicting the cell is uninfected when it is parasitized, and false positive would mean we are predicting the cell is parasitized when it is uninfected. Now, falsely treating patient who is uninfected could be a precaution, but falsely neglecting patient with malaria could cause tremendous harm to the patient, especially given that treating malaria after being untreated for a long time can actually be fatal to patient. We want low false negative rates as possible, which f1 score would represent, even if the accuracy of the model may be identical to other models.

Comparison of various techniques and their relative performance:

1. How do different techniques perform?
   1. Which one is performing relatively better?
   2. Pros and cons of different techniques
   3. Good to include a comparison table
2. Is there scope to improve the performance further?

Base model

Chart, line chart

Description automatically generatedTreemap chart

Description automatically generated with low confidence

Overall, the base model is solid. The test accuracy and validation accuracy are both relatively stable and it also has high f1-score. False negative and false positive cases are around the same. We observe quite high performance from base model, since the images in data set are quite uniform and it may not require very complicated model to accurately predict the class of the cell as parasitized or uninfected. We should explore more options.

Model 1

Chart, line chart

Description automatically generatedChart, treemap chart

Description automatically generated

A bit of drop in avg accuracy, but lower false negative cases. It should be considered as it provides us with a bit lower false negative case, since false negative case can be very fatal for a patient if Malaria has been left untreated even if we attempt to treat it later on. More options should be explored to see if we can achieve better performance, without sacrificing overall accuracy and rise in false positive rate, and drop in validation as epochs go on.

Model 2 using Batch normalization

Chart, line chart

Description automatically generatedChart, treemap chart

Description automatically generated

Compared to the base model. the accuracy has gone down, and also lower f1-score. We can also observe that the validation drops as epochs run more times. The model performs worse than the base model and should not be considered.

Model 3 Data Augmentation

Background pattern

Description automatically generatedThis is how the data was augmented. Horizontal flip was used, and it visually created pillar looking bars in the image. It may look like in human eyes that it is more distracting from cell’s feature and won’t help the model, but it needs to be compared in accuracy and f1 score after the model learns to see if the model actually performs better or worse.

Chart, line chart

Description automatically generatedChart, treemap chart

Description automatically generated

Contrary to human eye, the model actually performs better with horizontal flip data augmentation. It not only has the highest accuracy but also f1-score. The lowest case for false negative is proving that. Both validation and train accuracy are stable without a drop, and it is showing the best performance so far.

Chart, treemap chart

Description automatically generatedModel 4 Pre-trained model VGG16

Chart, line chart

Description automatically generated

Even though the VGG16 model is one of the most well-known and powerful model used around in industry, for our project, it is showing poor performance. The overall accuracy is low and does not increase as much as numbers of epochs run. The validation accuracy also fluctuates. While the false negative rate is lower than false positive rate (which is good for our project specifically), the f1-score shows that this model is creating more false-negative and false-positive cases overall, compared to the relatively simple base model. The model may be too complicated for our project. The reason could be due to the image’s simplicity and uniformity throughout the dataset.

Proposal for the Final Solution Design:

1. What model do you propose to be adopted?
   1. Based on the comparison, which is the best model for the problem?
   2. Think of the tradeoff between model performance and model interpretability
2. Why is this the best solution to adopt?
   1. Reason for choosing the best model
   2. How that solves the problem?

Text

Description automatically generated

As mentioned above, the data augmentation method used with CNN architecture showed the best performance overall in accuracy, f1-score, and low false negative cases for our project specific need. Our project specific need was that false negative case on Malaria and leaving a patient untreated for long periods of time could be fatal to patient, even when we attempt to treat Malaria, compared to false positive case where we treat normal person assuming one has Malaria. It is important to note that going through data augmentation of horizontal flip caused some image distortion of creating pillar looking bars in the background of the cell images, which at first sight may look like distraction to our human eyes, but it increased performance of our CNN model. A downside when it comes to interpretability could be that if a human doctor wanted the access to the particular cell images, after it goes through data augmentation, it may not look as distinguishable or helpful in human eye, but it brought up more model performance. Since data augmentation method has been chosen, a further exploration of different data augmentation is needed to observe and compare with method would bring out the most in model performance, such as HSV images or Gaussian blurring as it was done in Milestone 1.