Malaria Cell classifier

1. Introduction

In this project we need to build a classifier which predicts whether the cell is infected or not. This is a typical classification problem which can be solved using classification algorithms. In this project we will be using Convolution neural networks. Analysing cell images and confidently predicting whether the cell is infected or not is a big responsibility on the model, so we need to ensure that the model is at least 90 % confident on the test set. Our basic approach is to use Keras model and add Convolution layers as well as Dense layers to it and using all the standard optimizers and metrics ("adam" optimizer and "accuracy" metric). We are also building a streamlit app to show out model in the web browser. To Optimize our model further we will try to change the layer numbers(32, 64, 1024, etc.) and number of layers(1 conv layer, 2 conv layer, etc.) in the model and see the resultant model's performance and use the optimal layers and number of layers in our final model.

2. Problem Definition and Algorithm

2.1 Task Definition

The problem is to classify the uninfected and infected cells by looking at an image. The inputs to the model will be a PNG image and the output will be "Infected" or "not infected". This is a fascinating project as we have plenty different approaches of creating the model.

2.2 Algorithm Definition

In deep learning, a convolutional neural network is a class of deep neural network, most commonly applied to analyse visual imagery, so this perfectly fits our problem and that is why we are going to use this algorithm. CNN is a image recognition and processing that is specifically designed to process pixel data.

3. Experimental Evaluation

3.1 Methodology

Here the dependent variables are the labels("parasitized", "Uninfected") and the independent variables are the pixels of the image of size(50x50).[Note: We need to resize, convert to grayscale and normalize the images before passing it to our model]

1. Training set data: 19,297 images
2. Test set data: 8,267 images

We will be analysing our model on the basis of the loss metric and the accuracy metric. Also we will be generating heatmap and impose it on out image to see what areas affect the prediction. To optimize our model we have created a model which will run 27 times for the different values of layers and the number of layers. For example:

```
dense_layers = [0,1,2]
conv_layers = [1,2,3]
layer_sizes = [32,64,1024]
```

By comparing the model with different sizes we can pick the model with highest accuracy and lowest loss in both test and train set.

-Vishal Panchal

3.2 Results

Our model with the highest accuracy and minimal loss on both train and test set shows approximately 93 % accuracy. The streamlit app highlights the area that influenced the prediction.

3.3 Discussion

We can conclude that our model can predict whether the cell is infected or not with 93 % confidence/accuracy. We can tweak the model even further by changing the default optimizer and the default loss algorithm.

4. Related Work

Classification can also be achieved through logistic regression, though it is not that good as compared to Neural networks and CNN’s image recognition capability. The idea of using CNN is a lot better than that of using logistic regression as we all know that CNN is specially designed for processing pixel data.

5. Future Work

- Running the model without converting the images to grayscale.
- Tweaking the model’s default values to get even more accuracy.
- Building the model with Recurrent Neural Network algorithm.
- You can help to provide feedback on the model and suggest some future upgrades to the model

6. Conclusion

Our model can classify the infected cells by just passing a cell image. With this model in production the amount of time required to manually check the cell whether it is infected or not using an electron microscope can be reduced and also the cost and resources can be preserved and used for other fundamental analysis.

-Vishal Panchal