Malaria Infected Cell Classification

Malaria is a life-threatening disease. It's typically transmitted through the bite of an infected Anopheles mosquito. Infected mosquitoes carry the Plasmodium parasite.

When this mosquito bites you, the parasite is released into your bloodstream. Once the parasites are inside your body, they travel to the liver, where they mature. After several days, the mature parasites enter the bloodstream and begin to infect red blood cells. Within 48 to 72 hours, the parasites inside the red blood cells multiply, causing the infected cells to burst open.The parasites continue to infect red blood cells, resulting in symptoms that occur in cycles that last two to three days at a time.

Malaria is typically found in tropical and subtropical climates where the parasites can live. The World Health Organization (WHO)Trusted Source states that, in 2016, there were an estimated 216 million cases of malaria in 91 countries.

Deep Learning

Deep learning methods achieve impressive performance (even surpassing human-level performance) in several tasks over a wide spectrum of domains. However, in many domains like healthcare, reasoning the predictions of a model is vital. They are useful in evaluating the reliability and fairness of a model, and in effective data-driven decision making.

While being a generally black box approach, with modern techniques, deep learning methods can provide some explanations to their predictions. In the context of image classification, heat-maps on the image highlighting areas that highly influence the classification of the target can be generated.

Deep Learning Model Used: Convolutional Neural Network (CNN)

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. CNNs are powerful image processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks, often using machine vison that includes image and video recognition, along with recommender systems and natural language processing (NLP).
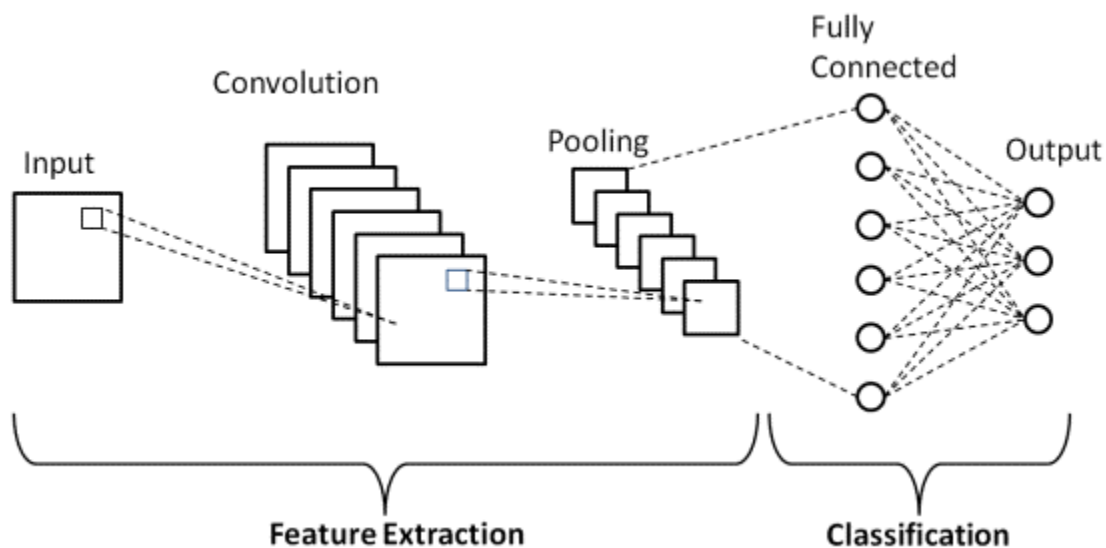
Dataset Collection

The dataset has been collected from the official website of National Institutes of health (National Library of Medicine).

https://lhncbc.nlm.nih.gov/LHC-downloads/downloads.html#malaria-datasets

Infected Malaria Cells

Uninfected Cells

Approach to the task

- The images are divided into training and validation sets using TensorFlow ImageDataGenerator

- Resizing of images into a 3D tensor with shape of width = 68, height = 68, channels = 3 shaped arrays, so that every image will have the same dimensions as each other.

- When performing image classification using CNN, it is recommended to rescale/normalize the images for less computational complexity.
  When using the image as it is and passing through a Deep Neural Network, the computation of high numeric values may become more complex. To reduce this, we can normalize the values to range from 0 to 1 by dividing it with 255 which is the highest pixel of an image.

Activation Functions Used

In a neural network, the activation function is responsible for transforming the summed weighted input from the node into the activation of the node or output for that input.

- Rectified Linear Unit (ReLU): The **rectified linear activation function** or **ReLU** for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is easier to train and often achieves better performance.
- Sigmoid: The sigmoid function is a special form of the logistic function and is usually denoted by σ(x) or sig(x). It is given by:

$$\sigma(x) = 1/(1+\exp(-x))$$

When the activation function for a neuron is a **sigmoid function**, it is a guarantee that the output of this unit will always be between 0 and 1. Also, as the sigmoid is a non-linear function, the output of this unit would be a non-linear function of the weighted sum of inputs. Such a neuron that employs a sigmoid function as an activation function is termed as a sigmoid unit.

Optimizer used

The choice of optimization algorithm for your deep learning model can mean the difference between good results in minutes, hours, and days.

- Adam: **Adam** is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data.
  According to Kingma et al., 2014, the method is "computationally efficient, has little memory requirement, invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data/parameters".

Loss Function Used

- Binary Cross Entropy: **Binary cross entropy** compares each of the predicted probabilities to actual class output which can be either 0 or 1. It then calculates the score that penalizes the probabilities based on the distance from the expected value. That means how close or far from the actual value.

Results

```
Epoch 1/6
1378/1378 [==============================] - 214s 155ms/step - loss: 0.4396 - accuracy: 0.8076 - val_loss: 0.2505 - val_accurac
y: 0.9265
Epoch 2/6
1378/1378 [==============================] - 47s 34ms/step - loss: 0.2346 - accuracy: 0.9226 - val_loss: 0.1970 - val_accuracy:
0.9258
Epoch 3/6
1378/1378 [==============================] - 49s 36ms/step - loss: 0.2084 - accuracy: 0.9358 - val_loss: 0.1816 - val_accuracy:
0.9383
Epoch 4/6
1378/1378 [==============================] - 46s 34ms/step - loss: 0.1909 - accuracy: 0.9423 - val_loss: 0.1719 - val_accuracy:
0.9385
Epoch 5/6
1378/1378 [==============================] - 52s 38ms/step - loss: 0.1771 - accuracy: 0.9464 - val_loss: 0.1673 - val_accuracy:
0.9359
Epoch 6/6
1378/1378 [==============================] - 45s 33ms/step - loss: 0.1586 - accuracy: 0.9511 - val_loss: 0.1742 - val_accuracy:
0.9327
```

We can see that the model is able to classify malaria-infected cells with an overall accuracy of ~ 96%

Streamlit

Streamlit is an open-source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc. With Streamlit, no callbacks are needed since widgets are treated as variables. Data caching simplifies and speeds up computation pipelines. Streamlit watches for changes on updates of the linked Git repository and the application will be deployed automatically in the shared link.

Streamlit apps are hosted in localhost:8501