

Machine Learning Engineer Nanodegree

Capstone Proposal

Abd Elrahman Elattar

22 December, 2018

Proposal

Domain Background:

AI (Artificial Intelligence) and machine learning is an important fields that helps in solving many problems like computer vision and self-driving cars. Recognizing objects in images and classifying it is an important task because it helps organizations to fasten their work to save its time.

In this project I have deep neural networks to train on 4 blood cell types. This can help in health care systems and it could be part of bigger systems.

The project is important because the average human adult has more than 5 liters (6 quarts) of blood in his or her body. Blood carries oxygen and nutrients to living cells and takes away their waste products. It also delivers immune cells to fight infections and contains platelets that can form a plug in a damaged blood vessel to prevent blood loss.

Through the circulatory system, blood adapts to the body's needs. When you are exercising, your heart pumps harder and faster to provide more blood and hence oxygen to your muscles. During an infection, the blood delivers more immune cells to the site of infection, where they accumulate to ward off harmful invaders.

All of these functions make blood a precious fluid. Each year in the USA, 30 million units of blood components are transfused to patients who need them. (Bethesda, 2005)

I have used this [paper](#) as a reference.

Problem Statement:

Blood cells have different types that it can make the doctors identify many diseases in many patients. So I want to make a deep learning model that can identify every blood cell and its type.

I have got the dataset from [kaggle](#) has 12,500 labeled images. I want to make a classifier that can take an image and predict its cell type.

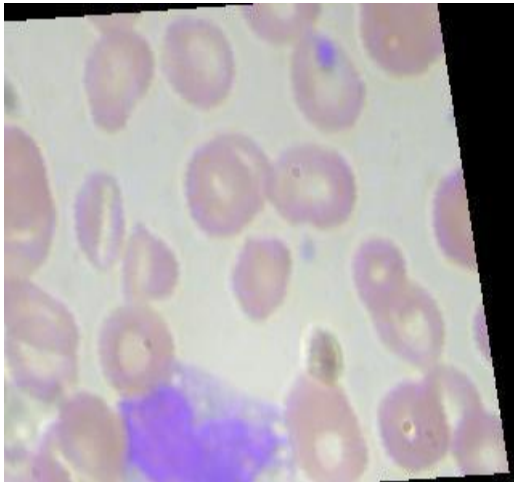
Datasets and Inputs:

I will use a dataset I have got from [kaggle](https://www.kaggle.com) which contains 12,500 blood cell images for the four types.

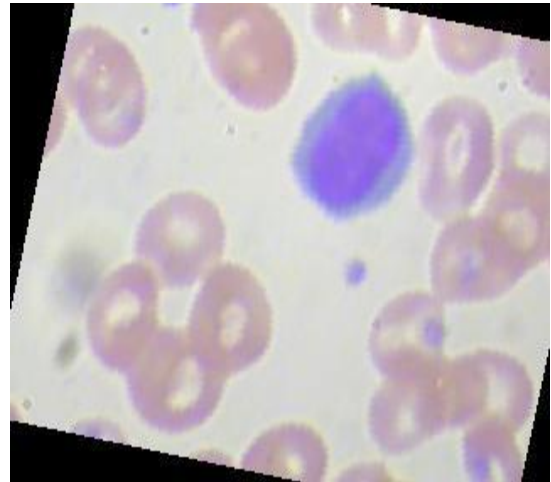
The content of dataset:

- Total number of the four types: 12,500.
- Train set: 8000.
- Valid set: 2500.
- Test set: 2000.
- Number of classes: 4
- Image size: 320X240

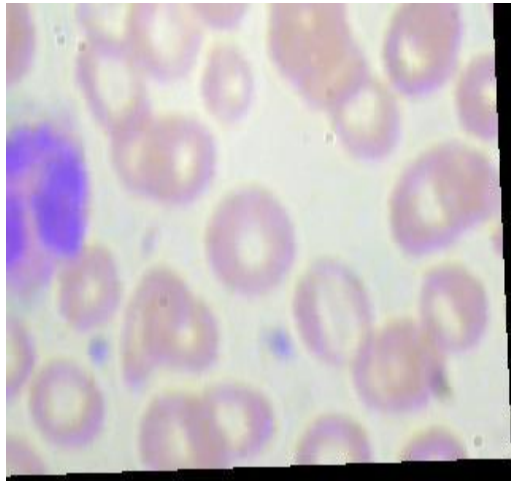
A preview for the four types:



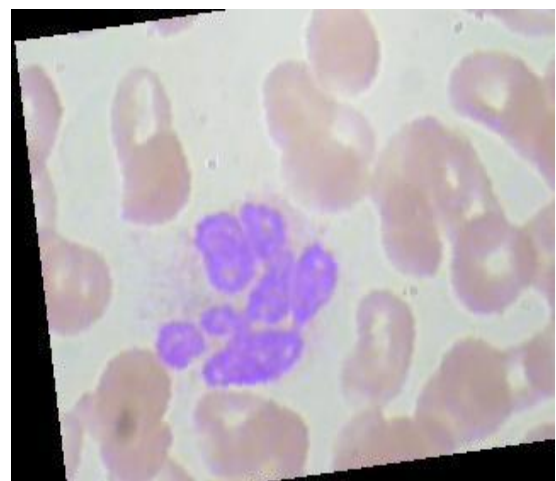
Monocyte



Lymphocyte



Eosinophil



Neutrophil

I have used a train files shape as (8000, 160, 160, 3) and a cell type shape (8000, 5) to input it to Keras CNN.

I will split the training set into training and validation set to maintain the balance of the dataset.

The data contains 2000 images for each type as a training set (total=8000) and 600 as a test for each type (total=2400).

Solution Statement:

The solution for this problem is to train a neural network model that can classify new inputted images to one of the four types.

I will use a Convolutional neural networks CNN to classify these images and to get a good accuracy.

Benchmark Model:

I will benchmark my model with a model in this [GitHub](#).

Model's structure:

```
model=Sequential()  
  
model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(60, 80, 3),  
strides=1))  
  
model.add(Conv2D(64, (3, 3), activation='relu'))  
  
model.add(MaxPooling2D(pool_size=(2, 2)))  
  
model.add(Dropout(0.25))  
model.add(Flatten())  
model.add(Dense(128, activation='relu'))  
model.add(Dropout(0.5))  
model.add(Dense(5, activation='softmax'))
```

The accuracy in this GitHub is 85%, I'm looking to approach to this accuracy.

Evaluation Metrics:

I will use accuracy test score on the test set to evaluate my model because it is appropriate if the dataset is balanced.

Project Design:

1. Import Dataset.
2. Preprocess Dataset. I will try to resize and normalize the images.
3. Making model architecture. I will try using CNN model from scratch and a pertained model like InceptionV3 or Resnet50.
4. Train the model.
5. Test the model.
6. Test the model on sample images.