WHITE BLOOD CELLS IMAGE CLASSIFICATION USING CNN AND SOBEL'S FILTER

A PREPRINT

Bala Krishna Chandana
chandana.b@husky.neu.edu

April 28, 2019

ABSTRACT

Detecting blood related diseases is one of the main tasks in todays world. Blood consists of three types of cells and they are Red Blood Cells, White Blood Cells and Platelets. Each type of cells have their own functionalities. Among these to detect and diagnose major diseases like Hemochromatosis, Hemolytic anemia, Hemophilia, etc., we need to identify White Blood Cells(WBCs) and its types. There are five major types of white blood cells. Automated methods to detect and classify white blood cell types have important medical applications. Here I am using, supervised machine learning technique Convolutional Neural Network(CNN) to classify the microscoipic images into one of the five WBC types. I have implemented CNN twice on the dataset. Once on the orginal datset and other I have used CNN on preprocessed dataset using Sobel filter. Sobel filter is a edge detection technique which is used in image processing and computer vision. I am using Sobel's filter to find out what is it's impact on CNN and how is it going to effect CNN. I used confusion matrix and accuracy to know the effect of sobel filter.

1 Introduction

White blood cells, also called leukocytes (leuko, meaning white), make up approximately one percent by volume of the cells in blood. The role of white blood cells is to identify and target pathogens, such as invading bacteria, viruses, and other foreign organisms. White blood cells are formed continually; some only live for hours or days, but some live for years. They have nuclei and do not contain hemoglobin. The different types of white blood cells are identified by their microscopic appearance after histologic staining, and each has a different specialized function. The two main groups, are the granulocytes, which include the neutrophils, eosinophils, and basophils, and the agranulocytes, which include the monocytes and lymphocytes.

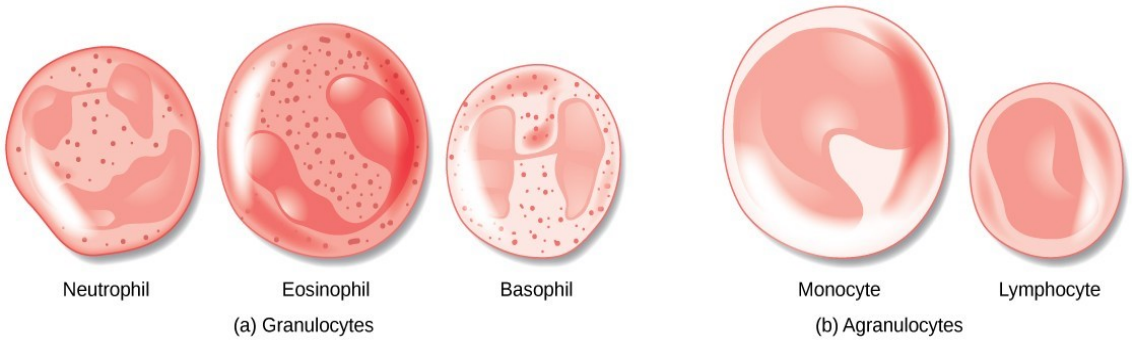


Figure 1: Whiteblood cells.

Usually, in hematology, microscopic analysis of peripheral blood smear results is time-consuming and costly process. Moreover, White blood cells often lead to misclassification and misidentification because they are inherently unstable.

At the same time, the statistical bias and inconsistencies of the hematologist further slow down the evaluation of the results. For all the above stated reasons, the development and use of computer-based systems instead of traditional methods will greatly contribute to the acceleration of the analysis and produce more accurate results.

This paper focuses on the classification of white blood cell images based on their statistical and geometrical feature information using deep learning technique Convolutional Neural Network(CNN) and Sobel's Edge Detection Technique. Based on the results we can know what is the effect of edge detection on CNN.

2 Dataset

The dataset contains 12,500 augmented images of blood cells (JPEG). There are approximately 3000 images for each of 4 different cell types grouped into 4 different folders (according to cell type). The cell types are Eosinophil, Lymphocyte, Monocyte, and Neutrophil. The folder 'dataset2-master' contains 2,500 augmented images as well as 4 additional subtype labels (JPEG + CSV). There are approximately 3,000 augmented images for each class of the 4 classes. There is also another folder 'dataset-master' which contains 410 images of blood cells with subtype labels and bounding boxes (JPEG + XML). These 410 images are original images(pre-augumentation). In this project I am using dataset2-master as the original images in other dataset have imbalanced class sizes. Our images are of size 320x420.

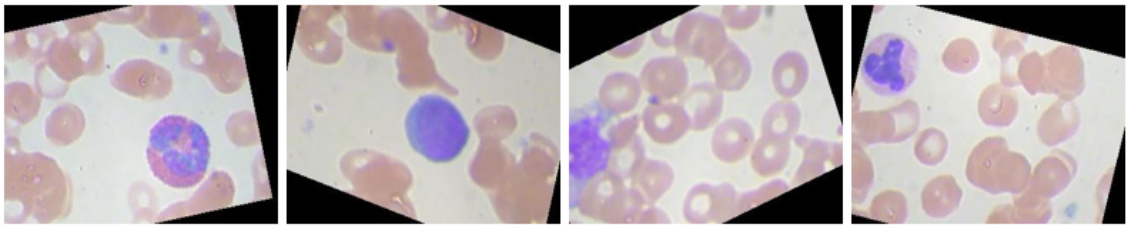


Figure 2: Whiteblood cells dataset images.

3 Technical Approach

The appraoch is to implement CNN twice on the dataset. Once on the orginal dataset and second time on preprocessed dataset using Sobel filter. By analyzing the results we can analyze Sobel filter's effect on CNN. We are going to use confusion matrix and accuracy to know the effect of sobel filter.

3.1 Convolutional neural network (CNN)

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, CNN have the ability to learn these filters/characteristics.

The architecture of a CNN is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area. CNN's are able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and re-usability of weights. In other words, the network can be trained to understand the sophistication of the image better.

3.1.1 Sobel Filter

Edge Detection is simply a case of trying to find the regions in an image where we have a sharp change in intensity or a sharp change in color, a high value indicates a steep change and a low value indicates a shallow change.

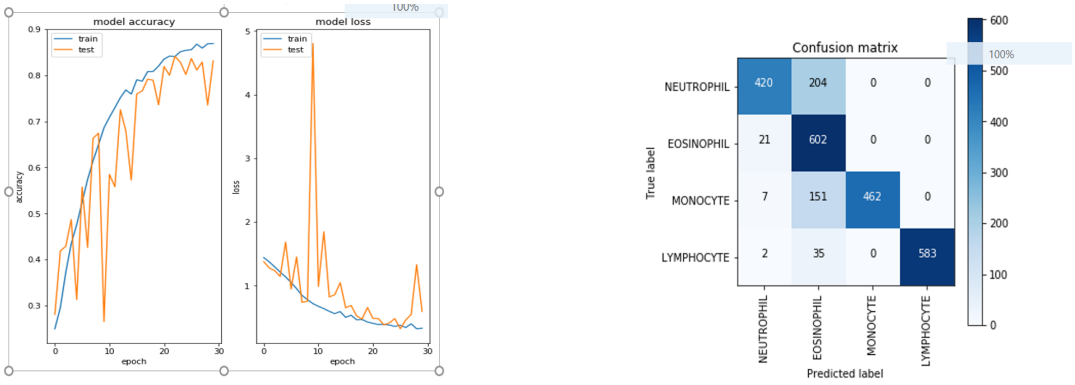
A very common operator for doing this is a Sobel Filter, which is an approximation to a derivative of an image. The Sobel operator, sometimes called the Sobel–Feldman operator or Sobel filter helps us to find out the amount of gradient difference by placing the gradient matrix over each pixel of our image.

4 Results

4.1 Using CNN on original data set

Used 2 Convolution layers each with a filter size of (3X3X32), 2 Dense layers, 2 Dropout layers and one Max Pooling layer. The model is trained using the individual images (60X80) of the WBC's and their corresponding one-hot encoded class labels. Plot 3a illustrates the accuracy of the CNN with respect to the number of epochs. We could observe from the graph that as the number of epochs increases, both the training and validation accuracy increases. It also represents the plot of Model accuracy and Model loss for the dataset. We could observe that the training and testing accuracy increases with the increase in the number of epochs and similarly, the model loss decreases as the number of epochs increases.

Figure 3b represents the confusion matrix obtained on the dataset of WBC images. As we can see from the figure, Lymphocytes were accurately predicted by the CNN classifier and there was some error in classifying true labels of Eosiophils. Around 32 percent of the Monocytes are misclassified as Neutrophils. This is because of the structure of Monocytes and Neutrophils are similar and the noise in the microscopic images led to the misclassification.



(a) Accuracy vs Epochs

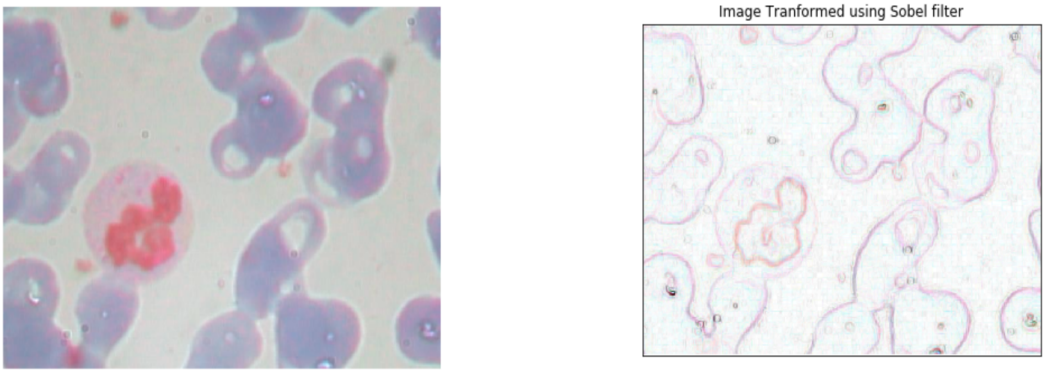
(b) Confusion Matrix

Figure 3: Results for CNN

4.2 Using CNN on Sobel pre-processed data set

4.2.1 Using Sobel filter directly on dataset

Using sobel filter the images are transormed as shown below. Fig 4a is the original image that we have in the dataset. Using the sobel filter the image is transformed as Fig 4b

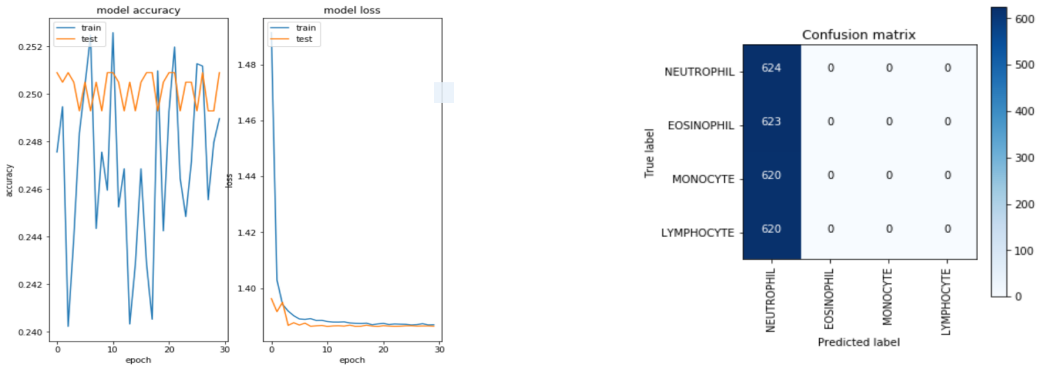


(a) Original Image

(b) Sobel Transformed Image

Figure 4: Sobel Transformation

So after transforming the images in the dataset to the above form, we run the CNN model on these images and got the below results.



(a) Accuracy vs Epochs

(b) Confusion Matrix

Figure 5: Results for CNN- 2

Plot 5a suggests that the accuracy is not linear with epochs and it is not consistent.

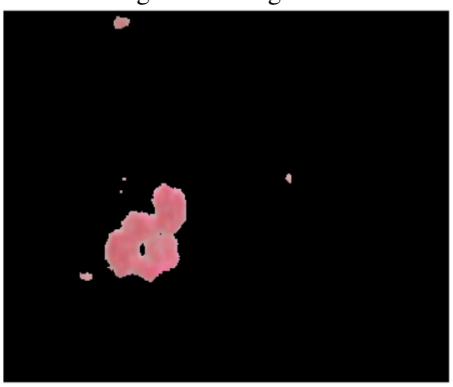
Figure 3b represents the confusion matrix obtained. From this we can see that the model predicted all the cells as NEUTROPHILS. This might be because when we detected edges of the images, we got the edges of all the other cells(WBCs) in the image. But RBCs are present in all the images. So the model might get thought all the images as same.

So before using sobel filter on the images, I am just getting the white blood cells from the images by using its color difference from red blood cells on the image. Which is done in the next section.

4.2.2 Pre-proccessing data before transforming using Sobel filter

To get only the white blood cells from the image I used OpenCVs cvtColor() method to convert our image to HSV Then defining the blue color range as wbcs in our image are in blue color. Then add these ranges as threshold to the HSV converted image to get only blue colors using OpenCv's inRange methods. Then mask the original image with this converted image.That way we only the image that looked like Fig 3b by converting image fig 3a. After getting this we applied sobel filter andd got the image that looked like Fig 6b

image.PNG image.PNG



(a) Preprocessed Image

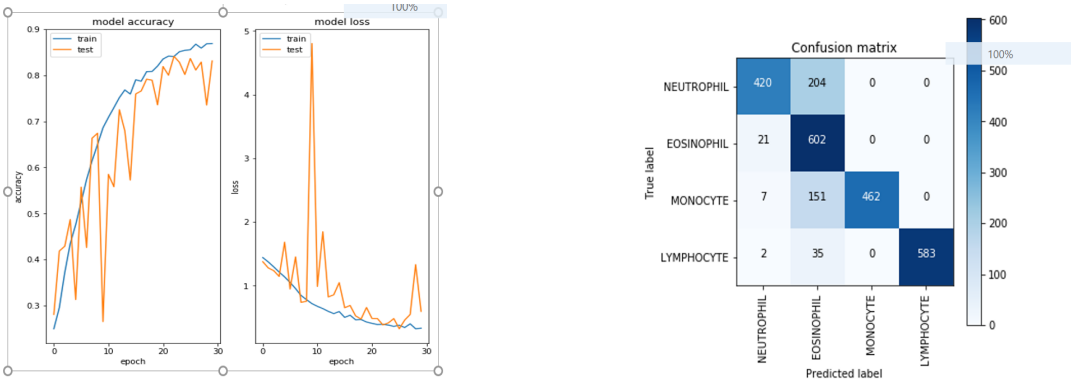
(b) Sobel Transformed Image

Figure 6: Sobel Transformation of pre-processed image

Plot 7a illustrates the accuracy of the CNN with respect to the number of epochs. We could observe from the graph that as the number of epochs increases, both the training and validation accuracy increases. It also represents the plot of Model accuracy and Model loss for the dataset. We could observe that the training and testing accuracy increases with the increase in the number of epochs and similarly, the model loss decreases as the number of epochs increases.

Figure 7b represents the confusion matrix obtained on the dataset of WBC images. As we can see from the figure, Lymphocytes were accurately predicted by the CNN classifier and there was some error in classifying true labels of

Eosiophils. Around 29 percent of the Monocytes are misclassified as Neutrophils. This is because of the structure of Monocytes and Neutrophils are similar and the noise in the microscopic images led to the misclassification.



(a) Accuracy vs Epochs

(b) Confusion Matrix

Figure 7: Results for CNN-3

4.3 Conclusion

For the original dataset I got an accuracy of 83 percent.For the pre-processed dataset using sobel filter direcly I got an accuracy of 25 percent. For the dataset where we only obtained whiteblood cells from images before transforming using sobel filter, I got an accuracy of 73 percent

The low accuracy for the second method might be because when we detected edges of the images, we got the edges of all the other cells(WBCs and Red blood cells) in the image. But Red Blood Cells are present in all the images. So the model might thought all the images as same. The confusion matrix confirms the same.

The low accuracy for third methods over the first one might be because of the of the similarities with in the types of cells. Like neutrophil and eusinophil are POLYNUCLEAR, similarly the other two types are MONONUCLEAR. So, since we detected the edges first using sobel filter in third method, the model might have predicted Neutrophil instead of eusinophil and vis-verca. It might have happened for other two types as well. The confusion matrix data also confirms the same. If we observe carefully, model predicted 152 NEUTROPHILS as EUSINOPHILS and 168 EUSINOPHILS as NEUTROPHILS.

So in the whole, we can say that Sobel's pre-processing doesn't help much in the image classification of white blood cells, infact it reduced the accuracy of the model.

References

[1] Kaggle dataset -https://www.kaggle.com/paultimothymooney/blood-cells

[2] Elen, Abdullah, and M. Kamil Turan. "Classifying White Blood Cells Using Machine Learning Algorithms." Uluslararas Mhendislik Aratrma ve Gelitirme Dergisi 11.1: 141-152.

[3] Bikhet, S. F., Darwish, A. M., Tolba, H. A. Shaheen, S. I. (2000). Segmentation and classification of white blood cells. IEEE International Conference on Acoustics, Speech, and Signal Processing, stanbul, pp. 2259-2261.

[4] Github-https://github.com/Shenggan/BCCDDataset.

[5] Medium - https://medium.com/datadriveninvestor/understanding-edge-detection-sobel-operator-2aada303b900

[6] Convolutional Neural Network - https://en.wikipedia.org/wiki/Convolutional neural network

[7] scikit-learn - https://scikit-learn.org/stable/documentation.html