

# RECOGNITION OF FOUR TYPES OF WHITE BLOOD CELLS IN PERIPHERAL BLOOD

Sravan Surendran

Student, Dept. of Dual Degree Computer Applications, Sree Narayana Guru Institute of Science and Technology  
 North Paravoor, Kerala, India

\*\*\*

**Abstract**-Recognition and inspection of white blood cells in peripheral blood can assist hematologists in diagnosing diseases like AIDS, leukemia, and blood cancer, making it one of the most salient steps in hematological procedures. This analysis can be accomplished by automatic and manual approaches. Manually scrutinizing blood cells by experts are tedious and susceptible to error. Therefore, an automatic system based on image processing techniques can help the hematologists in faster and efficient diagnosis. The field of Machine Learning, especially Convolutional Neural Networks (CNN) has shown promising prospects in this area. These algorithms can help in the analysis and segmentation of huge amount of data in very little time, without expert human intervention. This work aims to develop a deep learning model to handle the blood corpuscle classification problem, which is one among the foremost challenging problems in blood diagnosis. A CNN-based framework is built to automatically classify the blood cell images into subtypes of the cells (Neutrophils, Eosinophils, Lymphocytes, and Monocytes). The BCCD dataset is used for training the machine learning model.

**Key Words:** (convolutional neural network, machine learning, white blood cell detection, normalization, image preprocessing)

## 1. INTRODUCTION

The microscopic inspection of blood provides diagnostic information concerning patients' health status. The differential blood count inspection results reveal a good range of serious hematic pathologies. For example, the presence of infections, leukemia and certain specific types of cancers can be diagnosed based on the classification results and the white blood cell count. Experienced operators perform normal way for differential blood count. They use a microscope and count the share of every sort of cell that's counted within area of interest. This manual process of counting is obviously very tedious and slow. Furthermore, the classification and accuracy of the cell may depend on the operators, capabilities and experiences. Consequently, the need for an automated system of differential counting becomes inevitable. Recently, a number of different approaches have been proposed to implement a white blood cell recognition system based on image processing. White blood cell classification usually involves the following three stages: a white blood cell segmentation from an image, the

extraction of effective features, and a classifier design. To some extent, the performance of an automatic white blood cell classification system depends on a good segmentation algorithm to segment white blood cells from their background. We extract three sorts of characteristics from the segmented cell region below. These characteristics are fed into three different neural networks for the classification of 4 white blood corpuscle types. We extract three sorts of characteristics from the segmented cell region below. These characteristics are fed into three different neural networks for the classification of four white blood cell types.

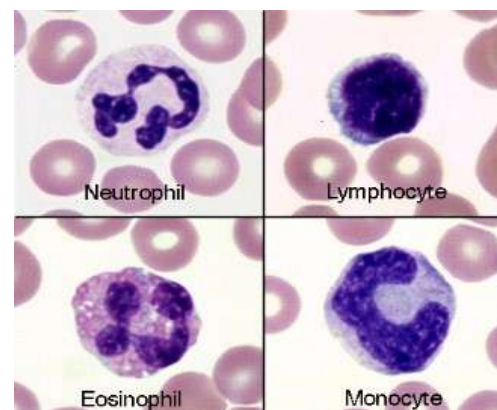


Fig-1 Four type of wbc cells

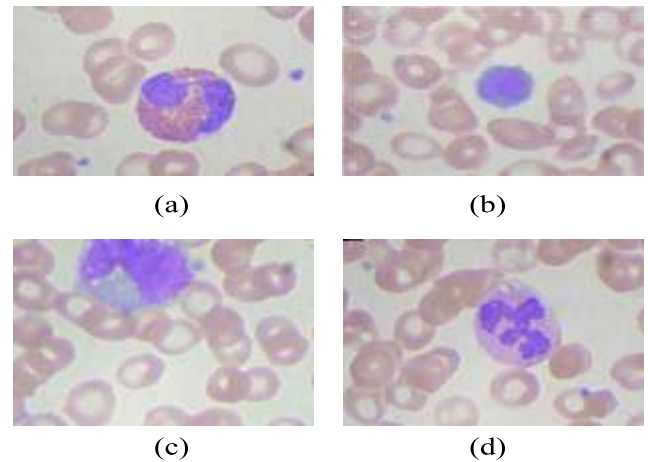
## 2. RELATED WORK

A review supported segmentation technique argues that conventional color based methods and thresholding methods are simple to sacrifice accuracy, whereas methods like region -growing offers high accuracy with high computation costs. Some methods work directly on the RGB color space, while others work directly on HSI or CMYK color space, referring to the color -based segmentation methods. In general, methods based on the S -component outperform those supported the RGB. By leveraging the CMYK color models, Putzu et al. (2013) attempts to create the feature vector. They determine that each one the opposite components except white blood cells have some yellow color them, while leukocytes show a good contrast in the CMYK color model's Y component. Young adopted four characteristics and a minimum distance classifier to classify 4 cell types. Wavelet transform coefficients and artificial neural networks employed by Sheik et al. to acknowledge white blood cells,

red blood cells, and platelets. Bikhet et al. selected some features and used minimum distance classifier to make an automatic arrangement that achieved a 91 percent correct classification rate for a 71 white blood cell database. Piuri and Scotti proposed an automatic classification and detection system supported 23 morphological characteristics and a neural classification system. An arrangement was proposed in supported own-cell and parametric characteristics. Nilufar et al. proposed a system of classification supported joint histogram-based characteristics and a vector support machine. Osowski et al. presented a genetic algorithm and a vector supporting machine for the popularity of blood cells within the bone marrow. Rezaatofghi et al. adopted morphological characteristics and textural characteristics extracted from local binary pattern (LBP) and then trained two types of neural networks for classification. Tabrizi et al. adopted the most component analysis for selection of features and used a neural network of learning vector quantization to classify 4 sorts of white blood cells. Ghosh et al. provided Naïve Bayes classifier with four statistically significant features to classify five sorts of white blood cells with an overall accuracy of 83.2 percent. Each approach has its own considerations for adopting features and classifiers of what kinds.

### 3. CONFIGURATION OF CNN

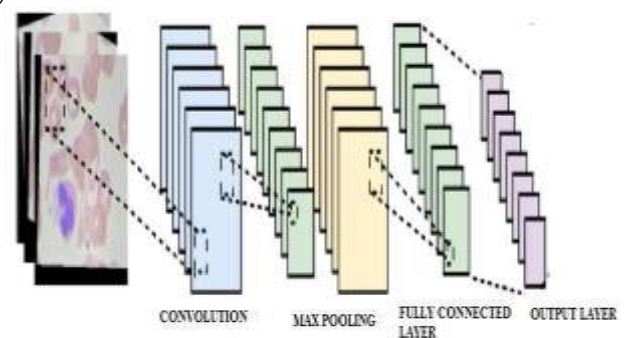
CNN is useful in classifying images and recognizing objects. It takes the raw pixels as input and produces an outcome indicating the probabilities that the input belongs to different classes. Instead of implementing the fully connected structure in each layer, CNN imposes two additional layers, convolution and pooling, which may significantly reduce the parameter magnitude. The convolutionary operation entitles the Convolution layer to extract the features from the input images. The Kernel modification generates features that have the implications of the variant, like object identification, image sharpening, edge detection etc. The pooling procedure is additionally mentioned as sub-sampling or down-sampling, which is intended to reduce the convolved characteristics produced by the convolution operator with the incentive to remain the significant information. There are various methods of pooling, such as maximum, average, summation, etc. In our method we use the maximize pooling. A single node is connected to all nodes in the previous layer in a fully connected layer. Moreover, quite one hidden layer may apply, and therefore the different classification operator could also be employed by the output layer.



**Fig-2** Database comprising 9957 cell images of (a) eosinophil, (b) lymphocyte, (c) monocyte, and (d) neutrophil.

### 3.1 Convolution layer

As the most important part of the convolutional neural network, the main way to calculate this layer is to use convolution windows with different sizes in order to perform convolution operations with the feature maps of the previous layer. Convolution windows of different sizes slide in sequence onto the feature map of the previous layer. The window size is usually  $3 \times 3$  or  $5 \times 5$ , and the number of weight parameters of the convolutional layer also changes accordingly. The values of the neurons on each feature map in the convolutional layer are convoluted through corresponding windows, and then the final result is obtained based on the excitation function used in the layer.



**Fig-3** Convolutional Neural Network

### 3.2 Max pooling

The calculation process of this layer is similar to the operation of the convolutional layer. The difference is that the sliding window of the lower sampling layer is usually  $2 \times 2$ , and the sliding step is 2. Therefore, this process will usually halved the feature map of the size of the previous layer, which to a large extent can greatly reduce the convolution weights of neural network parameters, the number of there are very good for the overall speed of the network training process to promote. At the same time, it

also enables the network to become more adaptive to the scale of the image changes.

### 3.3 Fully connected layer with softmax output

After the features generated by the RNN are merged with the features generated by the CNN, they are passed to the fully connected Softmax layer, the output of which is the probability distribution of all classes. In addition, we use the cross-entropy as a loss function to measure the difference between the actual output and the target output.

## 4. MODEL EVALUATION

### 4.1 Data set

The dataset consists of 9957 images with 4 different classes. The images are distributed unevenly between those classes and hence the model may predict some classes more accurately than other classes. I can populate the dataset with various image modifying techniques such as rotation, colour distortion or blurring the image. I will be training the model on the original dataset and will see the accuracy of the model. Then I will be adding more data and making each class even and check the model's accuracy.

### 4.2 Data preprocessing

Data Pre-processing refers to some artificial transformations to the raw dataset (including training, validation and testing datasets) in order to make the dataset more clean, more featureful, more learnable and in a uniform format. In order to improve the accuracy of the model and reduce over-fitting, we need to enhance the dataset. The data pre-processing is done before feeding the data to the data to the CNN model. In a convolutional neural network it is a fact that the performance of CNN is directly proportional to the amount of data used to train it, i.e. good pre-processing, always increases accuracy of the model. But on the other side, a bad pre-processing can also reduce the performance of the model.

### 4.3 Testing a CNN model

They are very successful in image recognition. The key part to know, which distinguishes CNN from traditional neural networks, is that the convolution operation. Having a picture at the input, CNN scans it repeatedly to seem surely features. This scanning (convolution) are often set with 2 main parameters: stride and padding type. As you can see in the picture below, the process of the first convolution gives us a set of new frames, shown here in the second column (layer). Each frame contains information about one feature and its presence in the scanned image. Resulting frame will have larger values in places where a feature is strongly visible and lower values where there are not any or little such features. Afterwards, the process is repeated for each of the obtained frames for a chosen number of times. In this project I chose a classic LeNet model which contains only two convolution layers.

## 5. CONCLUSION

In this paper, a classification model based on deep learning was implemented using Convolutional Neural Network to classify the image dataset into four WBCs — neutrophils, lymphocytes, eosinophils, and monocytes.

## 6. REFERENCES

- [1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," in Proc. ICLR, 2015, pp. 1-14.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770-778.
- [3] S. H. Rezaatofghi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," *Comput. Med. Imag. Graph.*, vol. 35, no. 4, pp. 333-343, 2011.
- [4] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 cell image classification with deep convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 2, pp. 416-428, Mar. 2017
- [5] Zamani F, Safabakhsh R. An unsupervised GVF snake approach for white blood cell segmentation based on nucleus. In: Proceedings of 8th international conference on signal processing. 2006.