The Implementation of Deep Learning for White Blood Cell Subtype Classification from Microscopic Images

Yustisia Amalia[1], Miftahul Khairoh[2], Arkha Rosyaria B.[3], Budi Santoso[4]
{yusti.amalia@unitomo.ac.id[1], miftahul.khairoh@unitomo.ac.id[2], arkha.rosyaria@unitomo.ac.id[3], budi.santoso@unitomo.ac.id[4]}

[1,2,3]Health Science Faculty, Indonesia
[4] Engineering Faculty, Dr.Soetomo University Surabaya, Indonesia

Abstract. This study aims to apply one of the artificial intelligence algorithms namely deep learning in the classification process of white blood cell subtypes. Classification of white blood cell subtypes is needed for identification of blood diseases such as leukemia, anemia, etc. The deep learning model used in this study is the LeNet architecture which is easy to implement. After going through the preprocessing of white blood cell microscopic images and classification of datasets according to the method, the classification accuracy was 94%..

Keywords: deep learning., white blood cell, classification.

1. Introduction

In the medical world, white blood cells or leukocytes have a very important role for humans. White blood cells are known as one of the human immune systems that are active against the harmful entry of foreign objects. In addition, the presence of white blood cells in the medical world is also used as an indication of several types of diseases. Besides the number of red blood cells or erythrocytes, white blood cells and platelets become a parameter in the diagnosis of diseases such as leukemia, anemia, cancer, and several other infectious diseases[1].
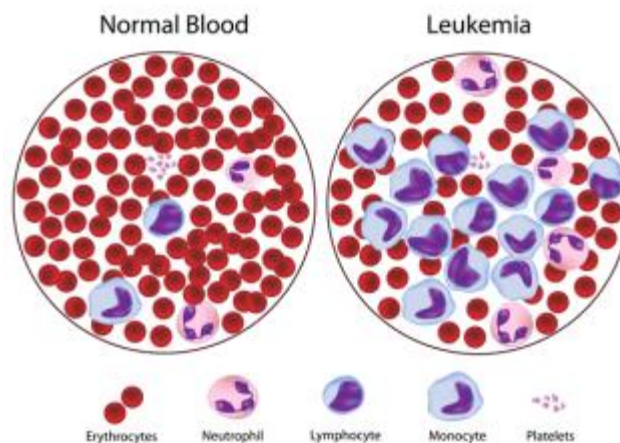

Fig. 1. More Lymphocytes on Leukemia patients (image source: MedGurus.org)

Diagnosis is done by making visual observations, so it requires great precision and patience by experts because mistakes are very possible.

Now with the rapid development of digital imaging technology, observations of blood cell parameters have been helped through medical image processing[2]. If previously the calculation of blood cells was done manually and directly, on the contrary in medical image processing this was done through a computational process that was fast and programmed. Therefore digital microscopic image processing occupies a very important position in hematology, a branch of medical science that focuses on the study of blood and hematopoietic tissue disease[3]. Many previous studies used microscopic image processing as a solution to medical problems such as screening TB bacteria[4], Chest Radiography[5], or telemedicine[6].

2. Backgrounds

This section will explain some important topics that underlie this research, including the theory and methods :

2.1 White Blood Cells

White blood cells or commonly abbreviated as WBC are cells that form blood components that function as part of the human body's weakness in fighting various infections. In a healthy adult human body, normally contains $4x10^9$ - $11x10^9$ white blood cells per liter of blood. Or on average there are 8000 white blood cells per mm cubic. In leukemia sufferers, this number has doubled. Therefore the number of leukocytes is one of the main parameters for the detection of leukemia. Leukocytes are not associated with one human organ but are able to move freely to the body and interact and capture cellular debris, foreign particles, or microorganisms. Leukocytes cannot divide or reproduce in their own way but are produced from pluripotent hematopoietic stem cells in the bone marrow.

Blood leukocytes are mainly divided into five categories, namely eosinophils, basophils, neutrophils, lymphocytes, and monocytes. The five types of leukocytes can be identified based on cell size, lobe nucleus, nucleus ratio with cytoplasm, cytoplasmic granules, and staining properties, etc.[7].

a. Neutrophils

Neutrophils are related to the body's defense against bacterial infections and other small inflammatory processes, and usually also provide the first response to bacterial infections; the activity and death of neutrophils in large amounts causes pus.

b. Eosinophils

Eosinophils are mainly associated with parasitic infections, thus increasing eosinophils indicates the number of parasites.

c. Basophils

Basophils are primarily responsible for allergic reactions and antigens by removing chemical histamine which causes inflammation.

d. Lymphocytes

Lymphocytes are more common in the lymph system. Blood has three types of lymphocytes:

B cells: B cells make antibodies that bind pathogens and destroy them. (B cells not only make antibodies that can bind pathogens but after an attack, some B cells will maintain their ability to produce antibodies as a 'memory' system service.)

T cells: CD4 + (helper) T cells coordinate resistance responses (which persist in HIV infection) and are important for holding intracellular bacteria. CD8 + (cytotoxic) can kill cells infected with the virus.

Natural killer cells: Natural killer cells (NK) can kill body cells that do not show a signal that they should not be killed because they have been infected with a virus or have become cancerous.

e. Monocytes

Monocytes divide the function of "vacuum cleaners" (phagocytosis) from neutrophils, but furthermore, he lives with an additional task: to give pathogens to T cells so that they can be memorized and killed, or can make antibody responses to guard.

2.2 Digital Image Processing

The development of digital image technology has penetrated the medical world, where at present most of the good visual observations for the process of identifying abnormalities in human organs can be done by computing machines with the help of digital cameras. Likewise with observations of microorganisms, as well as bacteria that cause human disease as can be identified through a computational process called digital image processing.

The image processing process begins by changing the value of the light intensity of each point from an image to a numerical matrix that represents the color values of Red, Green, and Blue from that point.

The next process is engineered against the value of the matrix until the desired results are obtained from digital image processing such as brightness or contrast enhancement, edge detection, color detection, size area, perimeter, and so on. These results are very useful for the greater importance of medical images that will be analyzed.

2.3 Machine Learning

A machine learning algorithm is an algorithm that is able to learn from data. A machine-learning system is trained rather than explicitly programmed. It's presented with many examples relevant to a task, and it finds statistical structure in these examples that eventually allows the system to come up with rules for automating the task[8].

For instance, to automate the task of tagging some pictures can be done by a machine-learning system with many examples of pictures already tagged by humans, and the system would learn statistical rules for associating specific pictures to specific tags.

Deep Learning

Deep learning is a special part of machine learning, a new view of learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations. The deep in deep learning isn't a reference to any kind of deeper understanding achieved by the approach, but it stands for this idea of successive layers of representations. How many layers contribute to a model of the data is called the depth of the

model. Other appropriate names for the field could have been layered representations learning and hierarchical representations learning. Modern deep learning often involves tens or even hundreds of successive layers of representations and they've all learned automatically from exposure to training data[9]. Meanwhile, other approaches to machine learning tend to focus on learning only one or two layers of representations of the data; hence, they're sometimes called shallow learning.
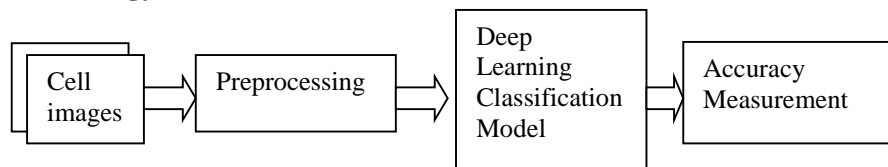
3. Methodology

Fig. 2. Block diagram of white blood cell subtype classification

The classification method of white blood cell subtype consists of 3 stages starting with preprocessing, which is a stage where the leukocyte microscopic image must be prepared to fit the system requirements. One process is to standardize all resolutions looking for leukocyte images. In addition, it is necessary to select one or two images that have too extreme differences such as brightness or contrast, and excessive noise.

The next stage is the classification process of leukocyte image data. This study uses deep learning based on CNN with the VGG network (VGGNet) architecture. This network is known for its simplicity because it only uses 3×3 convolutional layers that are stacked with one another in increasing depth. The reduced volume size is then handled with max pooling. Two fully connected layers, each with 4,096 nodes then followed by softmax classifier. More clearly can be seen in figure 3.
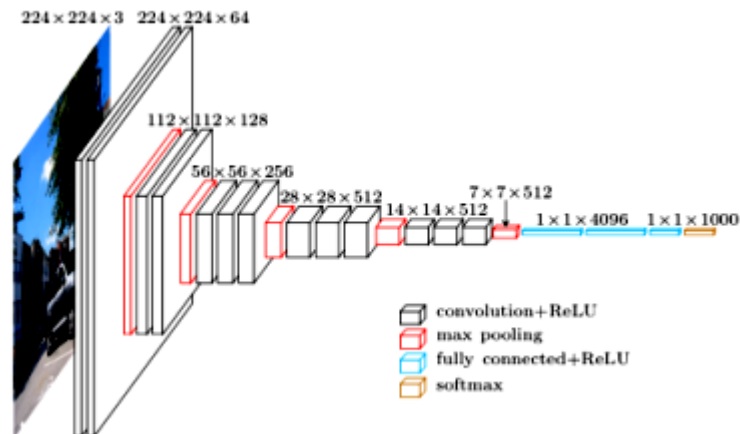

Fig. 3. A visualization of the VGG architecture (taken from cs.toronto.edu)

The last stage is the process of measuring accuracy. At this stage, the accuracy of the classification model used is calculated (the classification model based on CNN with the VGGNet architecture). Accuracy is obtained from the consistency of testing of datasets taken as data test.

The formula used in the calculation is:

$$Accuration = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} * 100\% \tag{1}$$

Where :

TP is True Positive, which is the amount of positive data that is correctly classified by the system.
TN is True Negative, which is the amount of negative data that is correctly classified by the system.
FN is False Negative, which is the amount of negative data but incorrectly classified by the system.
FP is False Positive, which is the number of positive data but incorrectly classified by the system

4. Dataset

The dataset used in this research experiment uses a dataset from Kaggle (www.kaggle.com) which held a contest to get the best method of classification of white blood cells. The dataset consists of 352 dead white blood cell image data with a resolution of 640 x 480 pixel. The pictures in the dataset have been labeled by the Pathologist and collected from the existing datasets.
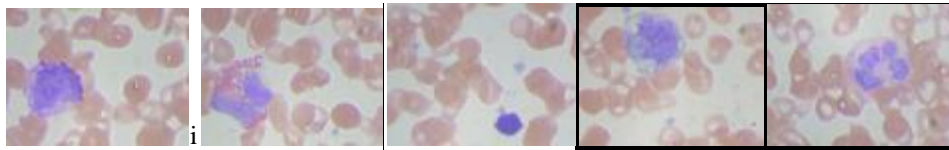

Fig. 4. test samples from datasets: Basophil, Eosinophil, Lymphocyte, Monocyte, and Neutrophil

5. The Results

The experiment is using some tools as follows:
-PC with Core I5 processor and 4GB RAM specifications
-Windows 7 64bit Operating System
-Python interpreter V3.4 with OpenCV 3, Hard, and TensorFlow
-Anaconda 3

a. Sample data calculation results

Based on the calculation process the sample data that has been labeled is obtained as figure 5.Since the dataset obtained is not balanced for all types of leukocyte images, it is necessary to add new data (augmented) by modifying existing sample data to obtain a balanced number of samples.

b. Accuracy calculation results based on the epoch

sAt the training process and accuracy calculations are obtained as follows:

```
from sklearn.metrics import accuracy_score

print('Predicting on test data')
y_pred = np.rint(model.predict(X_test_simple))

print(accuracy_score(y_test_simple, y_pred))
Predicting on test data
0.943661971831
```
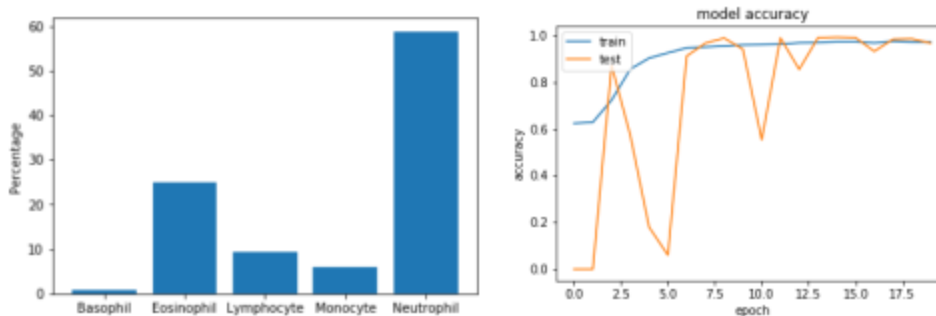

Fig. 5. Composition of data sample and model accuracy curve

The best accuracy is obtained at 0.94 or 94%. It is seen that the classification curve accuracy is increasing in proportion to the increase in the number of epochs. But this is followed by increasing time consumption needed during the training and testing process.

6. Conclusion

Based on the experiments conducted it can be concluded that the deep learning method for the classification process of white blood cell subtypes has succeeded well by obtaining an accuracy of 94%. This number is relatively high for the classification process compared to other machine learning methods. In addition, it was found that the classification process using deep learning requires a large and balanced number of datasets for each classified object.

References

[1] H. Cao, H. Liu, and E. Song, "A novel algorithm for segmentation of leukocytes in peripheral blood," *Biomed. Signal Process. Control*, vol. 45, pp. 10–21, 2018.

[2] E. Suryani, Wiharto, and N. Polvonov, "Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia," *Int. J. Comput. Sci. Netw. Solut.*, vol. 2, no. 6, 2014.

[3] I. E. Livieris, E. Pintelas, A. Kanavos, and P. Pintelas, "Identification of Blood Cell Subtypes from Images Using an Improved SSL Algorithm," *Biomed. J. Sci. Tech. Res.*, 2018.

[4] B. Santoso, A. Zaini, E. K. O. Pramunanto, H. Boedinugroho, and I. K. E. Purnama, "Penentuan Otomatis Posisi Fokus Citra Mikroskopis Bakteri Tuberkulosis Berbasis Nilai Entropi dan Fuzzy Logic Fuzzy Kendali Berbasis Logika."

[5] P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography : Automated Classification of Pulmonary Tuberculosis by Using Convolutional," vol. 0, no. 0, 2017.

[6] B. Santoso, "Color-Based Microscopic Image Steganography for Telemedicine Applications using Pixel Value Differencing Algorithm," in *IOP Conference Series:*

Journal of Physics: Conference Series 1175, 2019.

[7] Z.H.Al-Zubaydi, "White Blood Cells (WBCs) or Leukocytes," in *Medical Physiology*, 2014.

[8] S. Raschka, *Python Machine Learning*. 2015.

[9] A. Rosebrock, *Deep Learning for Computer Vision with Python*. 2017.