White Blood Cell Classification: A Comparison between VGG-16 and ResNet-50 Models

Supawit Vatathanavaro*, Suchat Tungjitnob†, Kitsuchart Pasupa‡
Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520
Thailand
Email: *supawit.vat@gmail.com, †suchat.tungjitnob@gmail.com, ‡kitsuchart@it.kmitl.ac.th

***Abstract*—White blood cell classification plays a significant role in helping a physician to diagnose disease. Using automated analyser machine can be easily analyse, fast, and accurate but the machine is very costly. Alternatively, this task can be manually perform by human who are expert in the field. However, it is very laborious. Machine learning and computer vision are applied to solve these limitations. In this study, two Convolutional Neural Networks–VGG-16 and ResNet-50–are employed to classify five types of white blood cell: Basophil, Eosinophil, Neutrophil, Lymphocyte, Monocyte. The results show that ResNet-50 is the best and can achieve 88.29 % accuracy.**

***Index Terms*—White Blood Cell, Deep learning, VGG-16, ResNet-50**

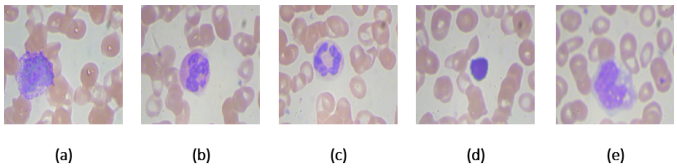
I. Introduction

White Blood Cells (WBC) classification is one of the important tasks that can assist medical doctors to diagnose disease [1]. There are two main methods for WBC classification task. The first one is to use automated blood analysers. Although this method can achieve very high accuracy, its cost and maintenance are incredibly high. The second method is human labour. Due to a similarity between the appearance of various types of WBC, it is very difficult to classify by the human eye. Hence, human should have an expertise in WBC classification. However, computer vision and machine learning are recently applied to solve these limitations. They can enable automatic approach with low cost. Many researchers applied several algorithms to WBC classification task, i.e., Support Vector Machine [2], Random Forest [3], Convolutional Neural Network (CNN) [4]. In [5], conventional machine learning are compared with deep learning algorithms. It is found that CNN is the best algorithm. Many deep learning architectures have been introduced to suit with different types of application, such as VGG-16 [6] and ResNet-50 [7]. Therefore, we aim to evaluate the performance of different architectures of CNN on WBC classification task. Here, we focus on VGG-16 and ResNet-50 that are distinctly different in terms of technique and depth. Both architectures can achieve very high accuracies on ImageNet. There are studies on deep learning with WBC [4], [8], [9], but, it is difficult to compare these algorithms together due to different experiment settings and datasets. Thus, we compare both architectures on a well-designed experimental framework on a combination of two different datasets in this work.

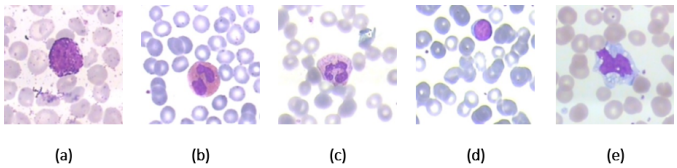
II. Methodology

A. Datasets

Two datasets are used as follows: (i) BCCD dataset–consists of 410 WBC images [10] and (ii) LISC dataset–consists of 400 WBC images [11]. Each dataset can be divided into five types of WBC namely Basophil, Eosinophil, Neutrophil, Lymphocyte, and Monocyte. Examples of WBC for each dataset are shown in Fig. 1.



(a) BCCD dataset.



(b) LISC dataset.

Fig. 1. Five types of WBC: (a) Basophil, (b) Eosinophil, (c) Neutrophil, (d) Lymphocyte, and (e) Monocyte.

B. Experimental Framework

Two datasets are combined together. The combined dataset is randomly stratified sampling into 80 % of a training set and 20 % of a test set. Because of a small number of sample in the training set, this can lead deep learning models to face over-fitting problem [12]. Therefore, we employed a technique called 'Image Augmentation' to generate additional images into the training set by rotating and flipping the WBC images. After the process, there are 1000 images in each type of WBC images that will be used as a training set.

In this work, we used pre-trained models that were trained from the ImageNet Dataset and fine-tuned their weight by our training set. Five-fold cross-validation was used to obtain the optimal models that were selected based on validation accuracy. The tuning parameter for both models is the number

of an epoch. After the optimal epoch is obtained, all training samples were used to create a model with the optimal setting. The model is evaluated by the test set.

III. Result and Discussion

As we mentioned in the previous section that we selected the optimal number of epoch based on average validation accuracy, the optimal number of epoch for the VGG-16 and the ResNet-50 is 44 and 47, respectively (as shown in Fig. 2). We report the average validation accuracy–based on the average validation accuracy across five folds–and accuracy that obtained from training and test sets as shown in Table I. The results show that the ResNet-50 model can achieve at 88.29 % of test accuracy that is better performance than the VGG-16 does. We further show the confusion matrices of the test set for both VGG-16 and ResNet-50 models in Table II and III. The results show that VGG-16 model can perform better in classifying Neutrophil while ResNet-50 model can perform better in classifying the others.

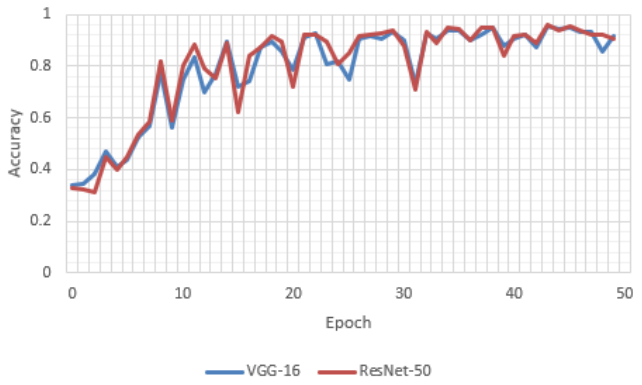


Fig. 2. Average validation accuracies across five-fold on training set by VGG-16 and ResNet-50 models.

TABLE I
The validation, training, and test accuracies of VGG-16 and ResNet-50 models.

Model	Validation (%)	Training (%)	Test (%)
VGG-16	95.30	86.74	72.07
ResNet-50	**97.26**	**99.62**	**88.29**

TABLE II
Confusion matrix of the test set by VGG-16 model.

Target	Predicted				
	Basophil	Eosinophil	Neutrophil	Lymphocyte	Monocyte
Basophil	4	1	0	1	2
Eosinophil	0	20	0	0	4
Neutrophil	0	0	17	0	1
Lymphocyte	1	1	3	7	1
Monocyte	1	13	0	1	33

IV. Conclusion

In this experiment, we adopted the VGG-16 and ResNet-50 architectures to WBC classification tasks. We evaluated the models on a combined dataset of two publicly available datasets. They consist of five types of WBC including Basophil, Eosinophil, Neutrophil, Lymphocyte, and Monocyte. The results show that ResNet-50 model can outperform VGG-16 model in this task. Refering to [6], network depth is a crucially important factor and the levels of features can be enriched due to the number of stacked layers [7]. If the depth of the VGG-16 model is increased, this can cause gradient vanishing problem that leads to higher training error [7]. On the other hand, the ResNet-50 model is deeper than the VGG-16 model, but it has an identity function that can preserve the gradient resulting in a more accurate model.

TABLE III
Confusion matrix of the test set by ResNet-50 model.

Target	Predicted				
	Basophil	Eosinophil	Neutrophil	Lymphocyte	Monocyte
Basophil	7	0	1	0	0
Eosinophil	0	21	0	0	3
Neutrophil	0	2	15	0	1
Lymphocyte	0	0	2	10	1
Monocyte	0	3	0	1	45

References

[1] M. Habibzadeh, A. Krzy zak, and T. Fevens, "Application of pattern recognition techniques for the analysis of thin blood smear images," *Journal of Medical Informatics and Technologies*, vol. 18, 2011.

[2] W. Tai, R. Hu, H. C. W. Hsiao, R. Chen, and J. J. P. Tsai, "Blood cell image classification based on hierarchical SVM," in *Proceedings of the IEEE International Symposium on Multimedia*, 2011, pp. 129–136.

[3] B. C. Ko, J. W. Gim, and J. Y. Nam, "Cell image classification based on ensemble features and random forest," *Electronics Letters*, vol. 47, no. 11, pp. 638–639, 2011.

[4] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," *Medical & Biological Engineering & Computing*, vol. 55, no. 8, pp. 1287–1301, 2017.

[5] W. Yu, J. Chang, C. Yang, L. Zhang, H. Shen, Y. Xia, and J. Sha, "Automatic classification of leukocytes using deep neural network," in *Proceedings of the IEEE 12th International Conference on ASIC (ASICON)*, 2017, pp. 1041–1044.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[8] M. Habibzadeh, A. Krzyżak, and T. Fevens, "White blood cell differential counts using convolutional neural networks for low resolution images," in *Proceedings of the 2013 International Conference on Artificial Intelligence and Soft Computing (ICAISC)*, 2013, pp. 263–274.

[9] M. J. Macawile, V. V. Quiones, A. Ballado Jr., J. D. Cruz, and M. V. Caya, "White blood cell classification and counting using convolutional neural network," in *Proceedings of the 3rd International Conference on Control and Robotics Engineering (ICCRE)*, 2018, pp. 259–263.

[10] Shenggan, "BCCD dataset," 2017, https://github.com/Shenggan/BCCD_Dataset, Accessed 01 August 2018.

[11] S. H. Rezatofighi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 333–343, 2011.

[12] K. Pasupa and W. Sunhem, "A comparison between shallow and deep architecture classifiers on small dataset," in *Proceeding of the 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2016, pp. 390–395.