Automatic Classification of Leukocytes Using Deep Neural Network

Wei Yu[1], Jing Chang[1], Cheng Yang[1], Limin Zhang[1], Han Shen[2], Yongquan Xia[2,*], Jin Sha[1,*]
[1]School of Electronic Science and Engineering, Nanjing University, Nanjing 210046, China
[2]Department of Clinical Laboratory, the Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing 210008, China
[*]Email: xyq2002xyq@aliyun.com, shajin@nju.edu.cn

***Abstract*—The classification of white blood cells is critical for the diagnosis of anemia, leukemia and many other hematologic diseases. Current approaches are mainly based on traditional machine learning methods, which take quite a noticeable time and the recognition error rate is relatively high especially for the rare kinds of leukocytes. In this paper, we develop an automatic cell recognition system by applying deep learning methods. The results demonstrate that the proposed system performs well in the leukocytes recognition task with less hardware limitations and higher accuracy compared to the traditional ones.**

***Index Terms*—Deep learning, leukocyte, cell classification**

I. Introduction

In the past few years, hematology test has been of great importance for the clinical diagnosis of many fatal diseases. In the blood routine test, the differential count of white blood cells is one of the key indicators which provide very useful information about the health conditions of patients. As the ratio of various kinds of leukocytes basically keeps stable in blood, the disturbance of proportion of white blood cells is a quite reliable symbol which suggests that the blood specimen comes from someone who is falling ill.

To get the proportion of leukocytes in the blood cell slides, it is necessary to classify the cells into several categories, namely, Monocytes, Lymphocytes, Basophils, Eosinophils, Atypical lymphocytes, Neutrophilic granulocytes, etc. The microscopic images of different types of leukocytes are shown in Fig. 1.

However, blood cell detection and classification is a highly demanding and labor-intensive work in medical laboratory. Currently, pathologists usually annotate the white blood cells manually. Considering the large data scale and the shortage of experienced doctors in clinical laboratory, it is not a satisfying solution.

In order to speed up the detection process and reduce the onerous manual work, several efforts have been made to automate the blood cell classification task.

In the 1970s, American and Japanese companies developed the earliest automatic white blood cells detection systems. However, they were not widely used due to the high price and low accuracy [1]. In [2], a method based on mathematical morphology (TSMM) was introduced. Later, P Sobrevilla et al. [3] developed an approach using fuzzy logic techniques. Huang D C and Hung K D used nucleus features to classify leukocytes with the help of K-means clustering algorithm.Leuko, a commercial leukemia diagnosis system, relied on naive Bayes classifiers to carry out cells classification task. Ushizima D M et al. [4] explored the possibility of using support vector machines (SVMs) in recognizing different types of leukocytes. Wang Shitong and Wang Min proposed a new detection algorithm based on fuzzy cellular neural networks [5]. However, these works were mostly evaluated with a limited amount of data or just focused on images taken from specified instruments, thus lacked good generalization capability.

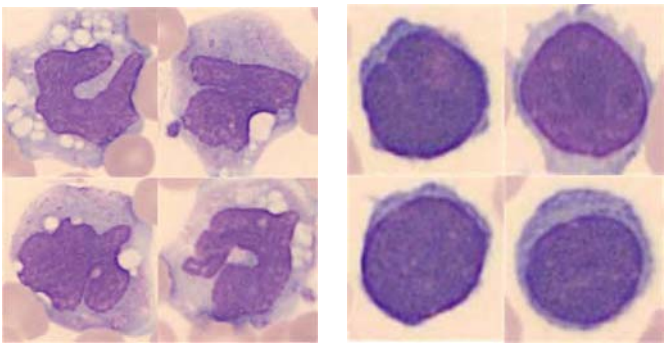
With the advancement of computational capabilities as well as the development of storage capacity for large scale data, deep learning has developed rapidly in recent years. Thus, the feasibility of automating laboratory analysis work using neural network has been researched in the recent literatures [6–9]. Recently, Andre Esteva et al. [10] used deep neural networks in skin cancer classification task and the performance was comparable to 21 skilled American dermatologists, which indicated the good potential of using deep learning algorithms in clinical analysis work.

In this paper, we present a automatic classification system that applies to the leukocytes classification mission. The system receives microscopic images of leukocytes as input and uses several state-of-the-art deep learning models to predict their categories. Then the system combines the results from different models and classifies the cells into the corresponding classes. Compared to the traditional methods, our new system can processes cell images with faster speed and higher prediction accuracy. It shows a bright prospect in the field of medical inspection.

II. Introduction of cell image database

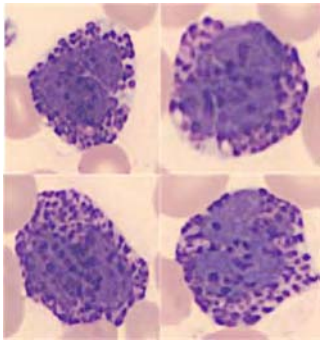
A new data set containing approximately 2000 microscopic images of 7 types of white blood cells is collected to train and evaluate our system. All images are provided by the Affiliated Drum Tower Hospital of Nanjing University Medical School and organized and labelled by experienced medical experts. Data masking work is applied to all the images to remove private information. Half of these images are used for training while the remaining images are set aside for the assessment of the generalization of our system.

978-1-5090-6625-4/17/$31.00 ©2017 IEEE

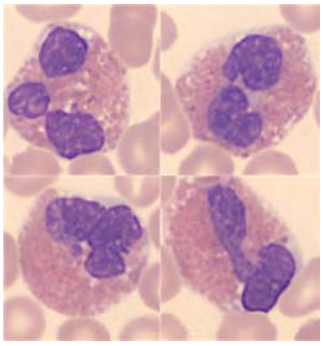


(a) Monocytes

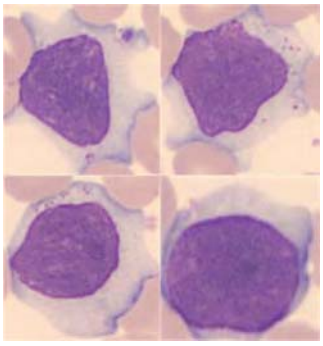
(b) Lymphocytes



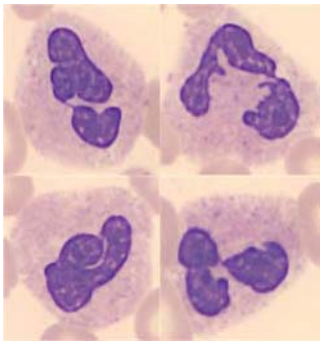
(c) Basophils



(d) Eosinophils



(e) Atypical lymphocytes



(f) Neutrophilic granulocytes

Fig. 1. Microscopic images of six typical types of white blood cells.

III. Method

Our system is mainly based on several state-of-the-art convolution neural networks.

A. Convolution Neural Network

Convolution Neural Network (CNN) is one of the basic network structures designed to perform machine learning tasks. As CNN is highly invariant to tilting, translation and scaling, it shows strong ability of adaptation in computer vision task. With the help of different convolution kernels, CNN can extract various local features from an image. Then these features are fed into a traditional neural network and an accurate output can be generated. The output result indicates the probability that a picture belongs to one or several certain categories. A typical architecture of CNN is shown in Fig. 2.

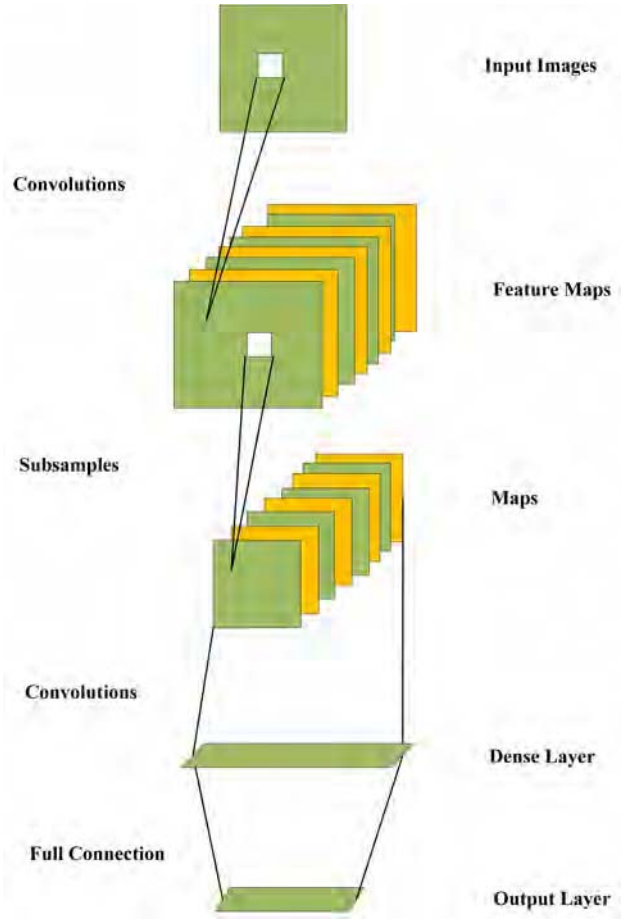


Fig. 2. A typical architecture of CNN.

B. Our Classification System Architecture

As single CNN model is easy to be effected by the data noise, the classification accuracy of one-model system may sometimes experience fluctuation. To get a higher and more stable accuracy, here we apply several different CNN models to the cell classification work and then combine their results to get a final result. The utilized network architectures include: ResNet50 [11]; Inception V3 [12]; VGG 16 [13]; VGG 19 [13] and Xception [14].

Considering the complex structures of these models and the limited data quantity, training these models from scratch using cell images will cost lots of time and computing resources while the result may not be very good. To tackle this problem, we use transfer learning in the work. All of these networks are pre-trained on data set from the ImageNet Large Scale Visual Recognition Challenge, which is composed of millions of images divided into about 1000 classes. Later we fine tune the models on our own training set. For each architecture, we discard its final layer and freeze the rest pre-trained weights during the following training process. Then, three fully-connected layers and two batch normalization layers are added to the top of each network respectively. Withal, the models are trained using our own training set one by one.

After that, we combine the classification results of different models with a voting mechanism to get the final result. Finally, statistical analysis is executed to evaluate our classification system on the test set. The architecture of our system is shown in Fig. 3.

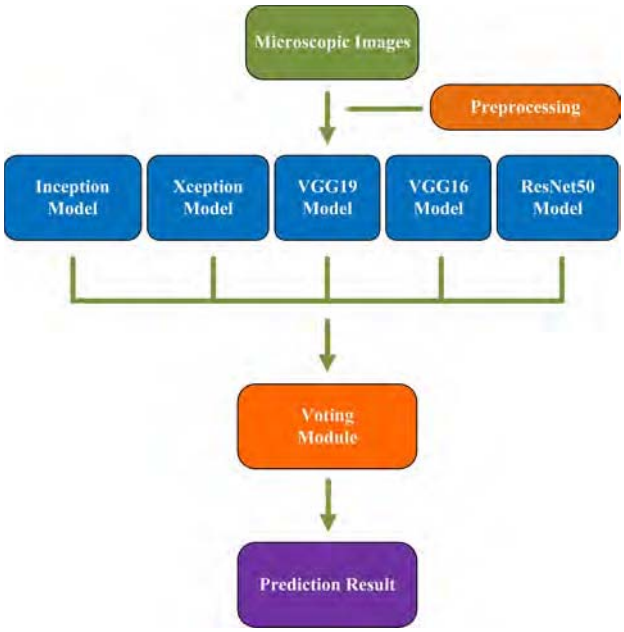


Fig. 3. Architecture of the classification system.

IV. Experimental results and comparisons

To evaluate the generalization capability of our system, we compare the prediction performance between our CNN model and several traditional models. All of these models are trained with the train images and evaluated in the test set. Table I demonstrates the performance of different models in the experiment. The result shows that our CNN model achieved an accuracy of about 88.5% in the classification task, which is higher than any other traditional methods.

TABLE I
PERFORMANCE COMPARISON BETWEEN DIFFERENT METHODS

Algorithm	Overall Accuracy
K-Nearest Neighbors (k=3)	82.90%
Support Vector Machine	84.68%
Logistic Regression	85.39%
Decision Tree	70.53%
CNN Model	88.5%

The corresponding confusion matrix of our system in the test set has been shown in Table II. From the summary statistics, several key indications can be calculated. In this experiment, we use sensitivity, accuracy, specifity and misdiagnosis rate to evaluate the performance of our CNN models in specific types of white blood cells. For each type, we classify the corresponding cells into four categories:true positive (TP), false positive (FP), true negative (TN) and false negative (FN). In this task, we define positive sample as the cell belonging to the specific type and negative sample as not belonging to. True positive (TP) is defined as the number of positive samples predicted by the model to be positive. When the system predict a positive sample as a negative sample, it is classified as false negative (FN). When a negative sample is classified as a positive one, it belongs to the false positive (FP) category. When the classification model indicates that a negative sample is negative, it is added to true negative (TN) category.

The accuracy is equivalent to the entirety of TP and TN separated by the aggregate number of leukocytes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{1}$$

The sensitivity is equal to the true equal rate.It is defined as follows.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2}$$

Specificity alludes to the true negative rate and is equivalent to the proportion of TN to the entirety of TN and FP.

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}. \tag{3}$$

Misdiagnosis rate is another key indicator. It can be calculated by the following formula.

$$\text{Misdiagnosis rate} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{4}$$

As can be seen from the Table III , our method not only performs well in the majority cell classes such as Lymphocytes and Neutrophilic segmented granulocytes, but also do well in the classify work of rare cells. For example, in the mission of classifying the rare Basophils and Eosinophils cells, our deep learning model achieves quite high accuracy with the help of a relatively small train data set. While for the most common white blood cells like Monocytes and Lymphocytes, the accuracies of our system are all over 95% and the misdiagnosis rates are close to 0% . In the experiment, even if blurred images are given, most of the time they can be classified correctly by our system. The experimental result clearly demonstrates the effectiveness of our proposed method.

V. Summary and future work

In this paper, we apply deep learning algorithms to the white blood cell classification task so the task can be executed in a completely automated way. The test result indicates that our model is able to classify different kinds of white blood cells in a short time with high accuracy. It is suggested that our model has a quite good generalization capability so it can be applied in clinic with less equipment constraints and less image sharpness limit. Further work will focus on accuracy improvement as well as parameters simplification to speed the classification. Furthermore, a bigger image database which contains not only leukocytes but also erythrocytes and platelets is now under construction so that our models can be trained to gain the ability of classifying all kinds of blood cells. It means that our system will be capable of automating the microscopic examination task of blood cell slides completely in the future.

TABLE II
CONFUSION MATRIX OF CLASSIFIER IN TESTING IMAGES

Actual class \ Predicted class	Monocyte	Lymphocyte	Basophil	Eosinophil	Atypical lymphocyte	Neutrophilic granulocyte	Neutrophilic stab granulocyte
Monocyte	20	1	0	0	21	0	1
Lymphocyte	0	78	0	0	50	0	0
Basophil	0	0	6	0	0	0	0
Eosinophil	0	0	0	12	3	0	3
Atypical lymphocyte	1	1	0	0	247	0	0
Neutrophilic granulocyte	3	3	0	1	37	668	9
Neutrophilic stab granulocyte	1	1	0	0	5	2	27

TABLE III
RESULT OF CLASSIFICATION IN TESTING IMAGES

WBCs	Accuracy	Sensitivity	Specifity	Misdiagnosis rate
Monocyte	96%	32%	100%	0%
Lymphocyte	95%	61%	99%	1%
Basophil	100%	100%	100%	0%
Eosinophil	99%	67%	100%	0%
Atypical lymphocyte	90%	99%	88%	12%
Neutrophilic granulocyte	95%	93%	100%	0%
Neutrophilic stab granulocyte	95%	93%	100%	0%

ACKNOWLEDGEMENT

The authors would like to express our sincerest gratitude to the doctors of the department of clinical laboratory at the Affiliated Drum Tower Hospital of Nanjing University Medical School, for their fully support and helpful suggestions in constructing our images database of white blood cells.

This work was jointly supported by the National Natural Science Foundation of China under Grant No. 61370040, 61376075, the project on the Industry foresight and common key technologies of Jiangsu Province BE2017153, and a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

REFERENCES

[1] P. K. Jr, "High-resolution leukocyte analyzers: retrospective and prospective," *Applied Optics*, vol. 26, no. 16, p. 3258, 1987.

[2] X. Yang, L. Luo, and Y. Wei, "Automatic classification system forleukocytes in human blood," *Chinese Journal of Computers*, vol. 17, no. 2, pp. 130–136, 1994.

[3] P. Sobrevilla, E. Montseny, and J. Keller, "White blood cell detection in bone marrow images," in *Fuzzy Information Processing Society, 1999. Nafips. International Conference of the North American*, 1999, pp. 403–407.

[4] D. M. Ushizima, A. C. Lorena, and A. de Carvalho, "Support vector machines applied to white blood cell recognition," in *International Conference on Hybrid Intelligent Systems*, 2005, p. 6 pp.

[5] W. Shitong and W. Min, "A new detection algorithm (nda) based on fuzzy cellular neural networks for white blood cell detection." *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, p. 5, 2006.

[6] S. J. Mckenna, I. W. Ricketts, A. Y. Cairns, and K. A. Hussein, "A comparison of neural network architectures for cervical cell classification," in *International Conference on Artificial Neural Networks*, 1993, pp. 105–109.

[7] H. A. Elsalamony, "Detection of some anaemia types in human blood smears using neural networks," *Measurement Science & Technology*, vol. 27, no. 8, p. 085401, 2016.

[8] S. Manik, L. M. Saini, and N. Vadera, "Counting and classification of white blood cell using artificial neural network (ann)," in *IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems*, 2017, pp. 1–5.

[9] J. Zhang, H. Hu, S. Chen, Y. Huang, and Q. Guan, "Cancer cells detection in phase-contrast microscopy images based on faster r-cnn," in *International Symposium on Computational Intelligence and Design*, 2017, pp. 363–367.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Corrigendum: Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[11] 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

[12] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567

[13] 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

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016. [Online]. Available: http://arxiv.org/abs/1610.02357