2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669587

TransMixNet: An Attention Based Double-Branch Model for White Blood Cell Classification and Its Training with the Fuzzified Training Data

1st Hua Chen
Institute of Artificial Intelligence
School of Computer Science, Wuhan University
Wuhan 430072, China

2nd Juan Liu
Institute of Artificial Intelligence
School of Computer Science, Wuhan University
Wuhan 430072, China
liujuan@whu.edu.cn

3rd Chunbing Hua
Institute of Artificial Intelligence
School of Computer Science, Wuhan University
Wuhan 430072, China

4th Zhiqun Zuo
Institute of Artificial Intelligence
School of Computer Science, Wuhan University
Wuhan 430072, China

5th Jing Feng
Institute of Artificial Intelligence
School of Computer Science, Wuhan University
Wuhan 430072, China

6th Baochuan Pang
Landing Artificial Intelligence Center for Pathological Diagnosis
Wuhan 430072, China

7th Di Xiao
Landing Artificial Intelligence Center for Pathological Diagnosis
Wuhan 430072, China

***Abstract*—White blood cells (WBCs) play a critical part in the human immune system, protecting human from bacteria and viruses. In general, WBCs can be divided into five categories: basophil, monocyte, eosinophil, neutrophil, and lymphocyte. The proportion of different kinds of WBCs is closely related to human health, thus analyzing the numbers and percentages of WBCs is usually included in the blood routine examination. Therefore, accurate identification of different types of WBCs has become an important step in the statistical analysis.**

However, the insufficiency and imbalance of WBC samples in medical computer vision domains is still a challenge to classify WBCs intelligently and accurately. This paper presents an attention based double-branch network, called TransMixNet, to build the classification model for WBCs recognition. During the model construction, we adopt transfer learning and data augmentation strategies to overcome the problems of insufficient and unbalanced samples. In addition, mixup strategy is used to model the domain relationship between different types of training samples to improve the generalization performance of the model. The evaluation experiment results show that our model is promising and applicable to clinical applications.

***Index Terms*—Double-Branch Network, Transfer Learning, Fuzzification, White Blood Cell Classification**

The work was supported by the Major Projects of Technological Innovation in Hubei Province (2019AEA170), the Frontier Projects of Wuhan for Application Foundation (2019010701011381), the Translational Medicine and Interdisciplinary Research Joint Fund of Zhongnan Hospital of Wuhan University (ZNJC201919).

I. Introduction

White blood cells (WBCs) play an important role in the defense of disease as well as foreign materials [1]. WBCs kill viruses via producing different kinds of antibodies. Thus, the number of various types of them can be regarded as a crucial indicator for a variety of diseases. The analysis of WBCs can reveal the state of patients' immune system [2]. WBCs are mainly divided into granular and non-granular cells according to their morphology and degree of granulation. The granular cells include neutrophil, eosinophil, and basophil, while the non-granular cells consist of monocyte and lymphocyte [3]. For a healthy person, neutrophils account for 50%-70%, eosinophils 1%-5%, basophils 0-1%, monocytes 2%-10%, and lymphocytes 20%-45% [4]. The relative quantity of different kinds of WBCs in the body may vary, and the ratio of subtypes of WBCs is closely related to many blood diseases such as leukemia, anemia, etc. Thus it is necessary to accurately classify and then analyze the proportion of different subtypes of WBCs.

During the 1950s, WBCs identification and counting were performed in a manual way. The manual method has the disadvantages of cumbersome, labor-intensive, time-consuming and strong subjectivity of results. Fortunately, the automated hematology analyzers which can count and identify the blood cells at high speed and accuracy has solved above problems

very well [5]. However, the hematology analyzers can neither meet the need for the morphological review of blood smears, nor digitally preserve the blood smears so as to support the retrospective analysis. Recently, rapid progress in digital microscopy and information technology have made it possible to generate the digital images of blood smears, which facilitate the retrospective study and the digital morphological analysis of blood smears [6]. What's more, with the blood cell images, many computer aided methods can be used for differentiating and analyzing WBCs efficiently and accurately.

Many researchers have developed computer aided methods to automatically classify different kinds of WBCs based on machine learning technologies. According to the adopted machine learning technologies, the existing methods can be divided into two categories: traditional machine learning based methods and deep learning based methods. The traditional machine learning based methods usually focus on extracting geometric, color, and texture features to represent WBCs and then building the classifiers by using traditional machine learning methods. For example, Falcon *et al.* [7] extracted contour-based and region-based shape features from manually segmented images of the nuclei, and used multi-layer perceptron, support vector machine (SVM), k-nearest neighbor, and C4.5 as classifiers to classify WBCs. Prinyakupt and Pluempitiwiriyawej [8] extracted the morphological features of the nuclei and cytoplasm, and then employed the linear and naïve Bayes classifier for the detection of WBCs. AL-Dulaimi *et al.* [9] proposed to construct L-moments features for nuclei shapes and coupled with the linear discriminant analysis, SVM, and classification tree to identify WBCs. Although these methods have achieved good performance, they greatly depend on the feature engineering. However, the extraction of features is hand-crafted, which requires a high level of expertise and takes a lot of time and effort.

Compared to the traditional machine learning techniques, deep learning methods can perform the end-to-end classification without the need of designing features in advance. Therefore many deep learning based methods have been proposed and achieved great progress. For example, Sharma and Kumar [10] built a convolutional neural network (CNN) classifier and obtained the 4-class classification accuracy of 98.14% on the BCCD data set. Baydilli and Atila [11] used the capsule networks to classify WBCs and achieved the accuracy of 96.86% on the public LISC data set [12]. With the observation that different network structures my have different concerns during the feature extraction and the classification, it has become a trend to combine different structures to build the classification models. Recent researches have illustrated that the combination of several deep neural networks can merit the prediction accuracy of the final models, such as Alexnet + Googlenet + DenseNet [13], Alexnet + Googlenet [14], canonical correlation analysis based CNN + LSTM (long short-term memory) [15], and so on. However, the above work of model combination seems to be just a try, and there is a lack of understanding whether the combination can give full play to the advantages of each model.

By deeply investigate the topologies of both ResNet and DenseNet within the HORNN framework, Chen *et al.* [16] have found that ResNet enables feature re-usage while DenseNet enables new features exploration, which are both important for learning good representations, hence they suggested the dual path networks to enjoy the benefits from both path topologies. Inspired by their work, we propose a double-branch network structure, TransMixNet, based on the combination of pre-trained ResNet and DenseNet. The proposed TransMixNet combines the latent features of the middle layers of ResNet and DenseNet that have been trained using the ImageNet to extract more effective features from the input. Meanwhile, we use an attention block in our model to facilitate learning from the dynamically-selected cell nucleus region of the WBC image. During the model training process, data augmentation is applied to alleviate the problem of insufficient and imbalanced data. In the meanwhile, we introduce the mixup operation to generate fuzzy training samples to further suppress the overfitting problem of the model. In order to evaluate the performance of TransMixNet, we conduct the comparing experiments on our indoor and the three public data sets.

II. Materials and Methods

A. Data Sets

We used four data sets in this work, one is the indoor data set collected by ourselves, and the other three are the public LISC [1], BCCD [2], and Raabin [3] data sets [17].

1) Indoor Data Set: We obtained blood samples from our cooperative medical institutions under the principle of medical ethics, and all samples do not involve any patients' personal information. We got the samples from 150 microscope slides of 150 normal subjects. The slides were smeared, Wright&Gimsa stained and scanned to get 150 digital slide images. After that, we applied our developed cell segmentation method to extract the cell images with size of 1280×1280 pixels in each slide image. Every cell image was labeled as one of five categories by at least two hematologists and the cell images with consistent labels from the hematologists were preserved. As a result, we collected 22,645 cell images, shown in Table I.

TABLE I
THE INFORMATION OF IMAGES IN DIFFERENT DATA SETS

Data Set	Image Number					Total	Pixel Size
	B[a]	M[b]	E[c]	N[d]	L[e]		
Indoor	224	968	539	10,469	10,445	22,645	1280×1280
LISC	53	48	39	50	52	242	720×576
BCCD	N/A	3,098	3,120	3,123	3,103	12,444	320×240
Raabin	301	795	1,066	8,891	3,461	14,514	575×575

[a]B: Basophil, [b]M: Monocyte, [c]E: Eosinophil, [d]N: Neutrophil, [e]L: Lymphocyte, N/A: Not Applicable.

[1]http://users.cecs.anu.edu.au/%7Ehrezatofighi/Data/Leukocyte%20Data.htm
[2]https://www.kaggle.com/paultimothymooney/blood-cells
[3]http://www.raabindata.ir/WBC/Cropped%5Fdouble%5Flabeled

2) Public Data Sets: The LISC, BCCD, and Raabin are publicly available WBC data sets containing multiple categories. The information of images in the three data sets is also shown in Table I. Note that the irrelevant background elements occupy a large portion of the images in the LISC and BCCD data sets, so that the images have low signal to noise ratios, which can negatively affect the classification performance. Therefore, for the LISC data set, the bounding box coordinates obtained by capturing the nonzero pixels in the masks given in the data set were utilized to crop WBCs in the images. For the BCCD data set, a total of 12,336 WBCs were extracted from the images by our cell segmentation method.

B. Architecture of TransMixNet

The architecture of our proposed TransMixNet is illustrated in Figure 1. ResNet and DenseNet are two kinds of typical CNNs, which have the characteristics of using the information of the previous layer so that the vanishing gradient problem may be alleviated when training deep networks. In addition, they also have the advantages of features re-usage and new features exploration respectively. Thereby, in order to benefit from these two aspects, we respectively chose the middle layers of ResNet and DenseNet without fully connected (FC) layers, called ResNet and DenseNet modules, as the beginnings of two branches in our network. At the end of ResNet and DenseNet modules, we used convolutional layers to replace their original FC layers to ensure the conformity of their output dimensions. The convolutional layer has 512 filters and a kernel size of 1×1 with stride of size 1 in this paper. As for the extracted features from two branches, we combined them together to get more information of WBC images. Since the nucleus contains most of the information about the cell, we introduced an attention block, Squeeze and Excitation (SE) block [18], so that the network could learn more features of the nucleus. Finally, we flattened the output and added two FC layers. To reduce the risks of overfitting, the dropout technique was also applied before the last FC layer. In the output layer, we exploited the softmax function as the activation function which yields the probabilities of all classes.

C. Data Processing and Model Training

1) Transfer Learning and Data Processing: Transfer learning is a machine learning technique that utilizes known knowledge learnt from one domain to solve different but related domain tasks. Transfer learning can solve the problem of insufficient knowledge due to limited data in the target domain to some extent, or can speed up the training process by fine-tuning parameters of a pre-trained model according to the target domain data rather than from scratch. ResNet50 [19] and DenseNet121 [20] have been trained on the ImageNet, so they have learnt enough features about the image texture, color, and so on. Therefore, we transferred the parameters of the middle layers of the two models into our network so that our model can focus on learning features specific to the WBC images.

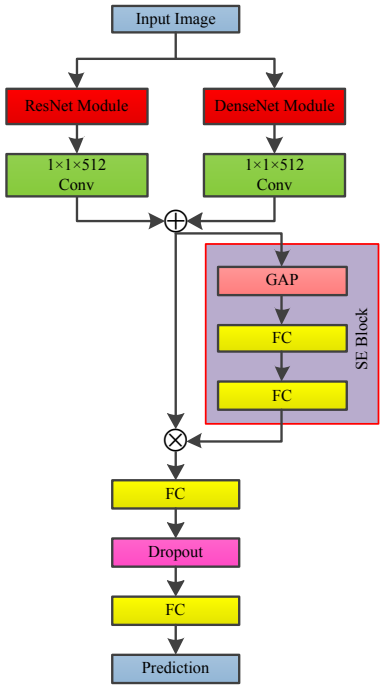

Fig. 1. Network architecture of classification. Conv: convolutional layer, GAP: global average pooling layer, FC: fully connected layer, $\oplus$: element-wise summation, $\otimes$: element-wise product.

Since the recommended input size of either ResNet or DenseNet is 224×224, we resized all the WBC images into 224×224 so as to fit the models. After that, the Indoor and LISC data sets were randomly divided into training, validation, and test sets at the ratio of 3: 1: 1 respectively. Given that the test sets have been provided in the BCCD and Raabin data sets, we randomly split the training data in the two data sets into training and validation sets in a 3:1 ratio respectively. The image numbers of different sets are listed in Table II.

TABLE II
THE IMAGE NUMBERS IN DIFFERENT SETS

Data Set	Total Number	Training Set	Validation Set	Test Set
Indoor	22,645	13,587	4,529	4,529
LISC	242	145	48	49
BCCD	12,336	7,404	2,467	2,465
Raabin	14,514	7,631	2,544	4,339

2) Data Augmentation: Generally speaking, deep neural networks can perform better with larger amount of data. However, due to the difficulty of data acquisition and the high cost of data annotation, the scale of labeled data that can be used for training model is generally small. In order to overcome the problem of insufficient data, data augmentation techniques have been proposed to increase the number of samples. In this study, we performed the data augmentation on the training sets in the Indoor, LISC, and Raabin data sets by randomly combining operations of flipping, rotating, translating, and changing brightness. Since the training set in the BCCD data set is already an augmented set, no additional

augmentation operations are performed here. Figure 2 shows the number of images in each category in the augmented training set in the four data sets.

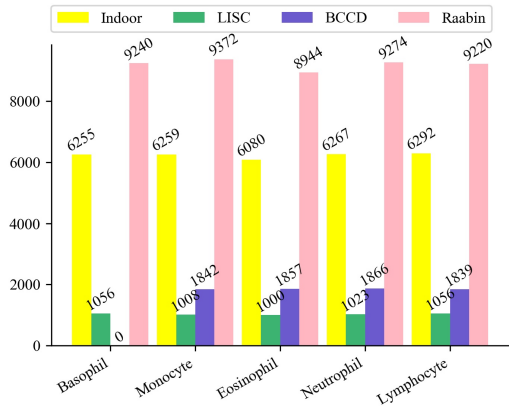

Fig. 2. Distribution of training set class sizes after data augmentation.

3) Implementation of Mixup Operation: Let a WBC image (x, y), where x is the pixel matrix, and y is the label. Now that x belongs to one of five categories, y can be represented by one-hot encoding. For example, $y = [1, 0, 0, 0, 0]$ means the cell belongs to basophil, while $y = [0, 1, 0, 0, 0]$ means the cell belongs to monocyte. Since medical and healthcare data can be subjective or fuzzy and people can usually make decisions based on imprecise information, we introduced "fuzzy" during the training process by performing mixup operation, which was originally proposed in [21], on the training data in this work. That is to say, we used the fuzzy rather than the original training samples to train our network.

For a training sample (x, y), we produced its fuzzified instance (x', y') by mixing up the information with another randomly selected training sample (x_i, y_i) by using the following formula:

$$x' = \lambda * x + (1 - \lambda) * x_i \tag{1}$$

$$y' = \lambda * y + (1 - \lambda) * y_i \tag{2}$$

where λ is the weight that satisfies the distribution of Beta (α, α), and $\alpha \in (0, +\infty)$ is a parameter.

For instance, let $y = [1, 0, 0, 0, 0]$, $y_i = [0, 1, 0, 0, 0]$, and $\lambda = 0.5$, we can obtain the fuzzy sample x' with label $y' = [0.5, 0.5, 0, 0, 0]$, which means that x' respectively belongs to basophil with 50% probability and monocyte with 50% probability according to its information. Figure 3 illustrates the generation of fuzzy image by mixing up basophil and monocyte images.

It is clear that mixup operation can be used to model domain relationships among different types of samples, leading to a linear interpolation of the associated labels, which is closer to the brain's mode of thinking and dicision-making.

4) Hyper-parameter Settings: In this study, our model was implemented on a 64-bit ubuntu 16.04 operating system with CUDA 9.0, Intel E5-2650 v4(2.20 GHz), RAM 256GB, GPU TITAN Xp(12 GB), and Python 3.5. The model settings

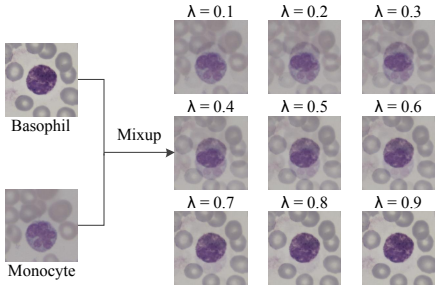

Fig. 3. An example of mixup operation.

referred to Keras with Tensorflow as backend supported by NVIDIA GPUs. During training, we selected cross entropy listed in (3) as the loss function and RAdam (Rectified Adam) [22] as the optimizer. The hyper-parameters were set as follows: epochs: 100, batch size: 16, dropout ratio: 0.5, initial learning rate: 1×10^{-5}.

$$L = -[y * log(\widetilde{y}) + (1 - y) * log(1 - \widetilde{y})] \tag{3}$$

where L means loss, y and $\widetilde{y}$ represent true label and predicted label respectively.

III. Experiments and Results

In this section, we first investigated the impact of fuzzification of training set by mixup operation on the model's performance. Then, we evaluated the effectiveness of combining ResNet and DenseNet modules, and SE block via the ablation study. Finally, we compared our proposed TransMixNet with the representative networks on WBC classification task.

A. Evaluation Criteria

The classification performance is assessed based on the following evaluation metrics: overall accuracy (OA), average recall (AR), average precision (AP), and average F1-score (AF1). In our five-class classification problem, positive samples are the samples in the considered class, and negative samples are the samples in the other four classes. The latter three evaluation metrics are calculated by calculating the corresponding average value of each independent class. The OA is defined as the ratio of the number of correctly classified samples to the total number of samples.

B. Investigation on Fuzzification of Training Samples

Since the parameter α determines λ in (1) and (2) and then determines the degree of fuzziness of training samples, we investigated the impact of α on the performance of our model by setting it from 0 to 1.0 with step 0.2. The performances with different α settings on the indoor test set are listed in Table III, from which we can see that the models trained with fuzzy samples usually achieve higher OA, AR, and AF1 than that trained with original samples ($\alpha = 0$). Moreover, it is not true that the larger α, the better performance of the trained model. According to Table III, we also found that the models trained with fuzzy samples obtained by setting α as 0.2 and 1.0

achieved better performance on the test set. We then further investigated the identification ability of the two models for different types of WBCs, shown in Table IV. In this table we can see that compared with α equal to 1.0, the model has slightly lower classification performance only for lymphocytes when α is set to 0.2. Thus, we set α as 0.2 to fuzzify the training samples to build our model.

TABLE III
THE PERFORMANCES (%) WITH DIFFERENT α SETTINGS ON OUR INDOOR TEST SET

α	OA	AP	AR	AF1
0	96.7	**94.8**	87.3	88.8
0.2	97.7	91.9	96.2	**93.9**
0.4	97.7	90.6	95.2	92.8
0.6	97.5	89.4	96.2	92.5
0.8	97.3	88.5	**97.0**	92.4
1.0	**97.8**	94.0	93.1	93.4

TABLE IV
THE PERFORMANCES (%) OF MODELS FOR EACH TYPE OF WBCS ON OUR INDOOR TEST SET

Type	α = 0.2				α = 1.0			
	Acc[a]	Pre[b]	Rec[c]	F1[d]	Acc	Pre	Rec	F1
Basophil	**100**	91.3	**100**	**95.5**	95.2	**93.0**	95.2	94.1
Monocyte	**85.6**	75.7	**85.6**	**80.4**	73.1	**86.4**	73.1	79.2
Eosinophil	**98.9**	**94.0**	**98.9**	**96.4**	**98.9**	**94.0**	**98.9**	**96.4**
Neutrophil	**99.7**	99.6	**99.7**	**99.7**	99.4	**99.7**	99.4	99.6
Lymphocyte	96.9	**98.7**	96.9	97.7	**98.7**	97.1	**98.7**	**97.9**

[a]Acc: Accuracy, [b]Pre: Precision, [c]Rec: Recall, [d]F1: F1-score.

C. Ablation Study on Each Component

As shown in Table V, when we removed the SE block, the OA, AR, and AF1 dropped, indicating that SE block has played a positive effect in enhancing features we extracted from the two branches. In addition, the model with fusing two branches achieved better performance than that with a single branch, which illustrated that our model inherited the advantages of ResNet and DenseNet modules, and made more effective use of the information of WBC images.

TABLE V
THE PERFORMANCES (%) OF ABLATION STUDY ON PROPOSED COMPONENTS ON OUR INDOOR TEST SET

TransMixNet			OA	AP	AR	AF1
ResNet Module	DenseNet Module	SE Block				
	✓	✓	97.4	91.7	93.4	92.4
✓		✓	97.5	89.3	96.0	92.4
✓	✓		97.5	**92.2**	93.4	92.6
✓	✓	✓	**97.7**	91.9	**96.2**	**93.9**

D. Comparison with Representative Methods

We compared TransMixNet with five state-of-the-art methods on our indoor and the public LISC, BCCD, and Raabin data sets to evaluate its performance. The hyper-parameter settings of all methods were the same. Table VI listed the comparison results of all methods on the test set of our indoor data, where each method were respectively trained with the original training data and the fuzzified trained data. From Table VI we can see that when training on fuzzified data, the performance of most methods has been improved, indicating that the model has stronger generalization ability after training on fuzzified data. Moreover, out TranMixNet achieved the best performance in our indoor data set. Then we further compared the methods on other three public data sets, all methods were trained with fuzzified data. The classification performances of the models on the test sets are shown in Table VII. In this table, our proposed TransMixNet still achieved the best overall classification results, demonstrating that our method is effective and has great potential in practical applications of the WBCs classification task.

TABLE VI
THE COMPARING RESULTS (%) OF DIFFERENT METHODS FOR ORIGINAL DATA AND FUZZIFIED DATA ON OUR INDOOR TEST SET

Model	Training Set	OA	AP	AR	AF1
ResNet [19]	original	96.5	89.2	87.7	88.4
	fuzzified	96.6	86.3	91.4	88.6
DenseNet [20]	original	96.5	89.1	90.3	89.4
	fuzzified	96.8	88.1	93.2	90.5
Inception v3 [23]	original	89.4	70.6	68.2	69.1
	fuzzified	88.7	73.3	63.6	67.2
Jiang [24]	original	97.1	90.9	90.7	90.7
	fuzzified	97.4	91.1	92.3	91.6
Sharma [25]	original	95.7	85.6	82.1	83.2
	fuzzified	95.8	89.0	79.8	83.5
TransMixNet	original	96.7	**94.8**	87.3	88.8
	fuzzified	**97.7**	91.9	**96.2**	**93.9**

TABLE VII
THE COMPARING RESULTS (%) OF DIFFERENT METHODS ON THE LISC, BCCD, AND RAABIN TEST SETS

Data Set	Model	OA	AP	AR	AF1
LISC	ResNet [19]	93.9	95.5	92.5	93.0
	DenseNet [20]	**98.0**	98.3	**97.5**	97.8
	Inception v3 [23]	75.5	74.9	74.6	73.7
	Jiang [24]	**98.0**	**98.5**	**97.5**	**97.9**
	Sharma [25]	95.9	96.8	95.0	95.5
	TransMixNet	**98.0**	98.3	**97.5**	97.8
BCCD	ResNet [19]	84.7	87.1	84.7	85.2
	DenseNet [20]	87.1	89.4	87.2	87.5
	Inception v3 [23]	62.8	67.7	62.8	63.5
	Jiang [24]	86.8	89.3	86.8	87.1
	Sharma [25]	87.0	89.1	87.0	87.3
	TransMixNet	**88.0**	**90.1**	**88.0**	**88.3**
Raabin	ResNet [19]	96.4	92.9	96.1	94.3
	DenseNet [20]	97.1	94.0	97.1	95.4
	Inception v3 [23]	89.6	78.4	88.3	82.5
	Jiang [24]	96.1	91.7	97.0	94.0
	Sharma [25]	96.0	92.6	95.1	93.5
	TransMixNet	**98.6**	**97.0**	**98.7**	**97.8**

IV. CONCLUSION AND DISCUSSION

In this paper, we proposed a novel deep learning framework, TransMixNet, for WBCs classification. The proposed TransMixNet integrated the advantages of ResNet and DenseNet modules, while benefiting from the guidance of attention

block, which could extract more effective information from WBC images. In order to solve the problem of insufficient and imbalanced data, transfer learning and data augmentation strategies were applied. Moreover, to avert the problem of overfitting, in addition to using dropout operation, we also proposed to use fuzzified instead of the original training samples to train our model. Comparing results with state-of-the-art networks on our indoor and the three public data sets have shown the excellent overall performance of our model. The investigation on the fuzzification also illustrates that it is a feasible candidate solution to solve the problem of overfitting.

However, we also noticed that our model is not as good at recognizing monocytes as other types of WBCs. From Table IV we can see that all measurements for monocytes are much lower than those for other types of cells. By further investigating the confusion matrix in Figure 4, we found that many monocytes were misclassified as lymphocytes. At the same time, the number of lymphocytes that were wrongly classified as monocytes is more than that of other types of WBCs. Such results show that our model needs to enhance the ability to differentiate monocytes from lymphocytes. We also consider extending our work from the current five categories of WBCs to more classes in the future. In addition, the model should be further optimized to reduce training time to better meet the needs of clinical applications in future study.

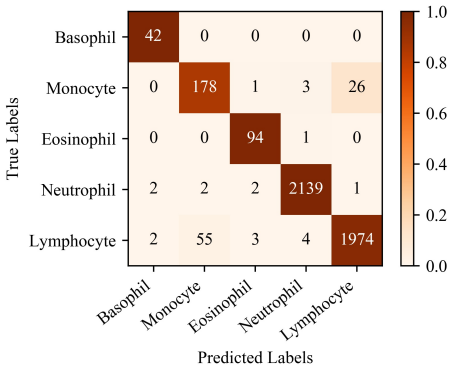

Fig. 4. Confusion matrix of classification results on our indoor test set.

REFERENCES

[1] S. T. Mohamed, H. M. Ebeid, A. E. Hassanien, and M. F. Tolba, "Optimized feed forward neural network for microscopic white blood cell images classification," in *Proc. International Conference on Advanced Machine Learning Technologies and Applications (AMLTA)*, Cairo, Egypt, Mar. 2019, pp. 758–767.

[2] M. Saraswat and K. Arya, "Automated microscopic image analysis for leukocytes identification: A survey," *Micron*, vol. 65, pp. 20–33, Oct. 2014.

[3] L. Putzu and C. D. Ruberto, "White blood cells identification and counting from microscopic blood image," in *Proc. International Work-Conference on Bioinformatics and Biomedical Engineering*, Granada, Spain, Mar. 2013, pp. 99–106.

[4] R. D. Labati, V. Piuri, and F. Scotti, "All-IDB: The acute lymphoblastic leukemia image database for image processing," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 2011, pp. 2045–2048.

[5] G. Chhabra, "Automated hematology analyzers: Recent trends and applications," *Journal of Laboratory Physicians*, vol. 10, pp. 015–016, Jan. 2018.

[6] H. E. Achi and J. D. Khoury, "Artificial intelligence and digital microscopy applications in diagnostic hematopathology," *Cancers*, vol. 12, pp. 797–811, Apr. 2020.

[7] A. Falcon, A. Taboada-Crispi, M. Orozco-Monteagudo, M. Aliosha-Perez, and H. Sahli, "Classification of white blood cells using morphometric features of nucleus," in *Proc. Cuba-Flanders Workshop on Machine Learning and Knowledge Discovery*, Feb. 2010, pp. 1–9.

[8] J. Prinyakupt and C. Pluempitiwiriyawej, "Segmentation of white blood cells and comparison of cell morphology by linear and naïve bayes classifiers," *Biomedical Engineering Online*, vol. 14, pp. 1–19, Jun. 2015.

[9] K. Al-Dulaimi, K. Nguyen, J. Banks, V. Chandran, and I. Tomeo-Reyes, "Classification of white blood cells using L-moments invariant features of nuclei shape," in *Proc. International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Auckland, New Zealand, Nov. 2018, pp. 1–6.

[10] G. Sharma and R. Kumar, "Classifying white blood cells in blood smear images using a convolutional neural network," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 825–829, Jul. 2019.

[11] Y. Y. Baydilli and Ü. Atila, "Classification of white blood cells using capsule networks," *Computerized Medical Imaging and Graphics*, vol. 80, Mar. 2020.

[12] 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, pp. 333–343, Jun. 2011.

[13] M. Toğaçar, B. Ergen, and Z. Cömert, "Classification of white blood cells using deep features obtained from convolutional neural network models based on the combination of feature selection methods," *Applied Soft Computing Journal*, vol. 97, Dec. 2020.

[14] A. Çınar and S. A. Tuncer, "Classification of lymphocytes, monocytes, eosinophils, and neutrophils on white blood cells using hybrid alexnet-googlenet-SVM," *SN Applied Sciences*, vol. 3, pp. 503–513, Apr. 2021.

[15] A. M. Patil, M. D. Patil, and G. K. Birajdar, "White blood cells image classification using deep learning with canonical correlation analysis," *IRBM*, 2020.

[16] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," in *Proc. International Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 4470–4478.

[17] Z. M. Kouzehkanan, S. Saghari, E. Tavakoli, P. Rostami, M. Abaszadeh, F. Mirzadeh, E. S. Satlsar, M. Gheidishahran, F. Gorgi, S. Mohammadi, and R. Hosseini, "Raabin-WBC: A large free access dataset of white blood cells from normal peripheral blood," *bioRxiv*, May 2021.

[18] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 2011–2023, Aug. 2020.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[20] G. Huang, Z. Liu, L. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.

[21] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[22] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," *arXiv preprint arXiv:1908.03265*, 2019.

[23] W. Yu, J. Chang, C. Yang, L. Zhang, H. Shen, Y. Xia, and J. Sha, "Automatic classification of leukocytes using deep neural network," in *Proc. International Conference on ASIC (ASICON)*, Guiyang, China, Oct. 2017, pp. 1041–1044.

[24] M. Jiang, L. Cheng, F. Qin, L. Du, and M. Zhang, "White blood cells classification with deep convolutional neural networks," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, p. 1857006, Feb. 2018.

[25] M. Sharma, A. Bhave, and R. Janghel, "White blood cell classification using convolutional neural network," in *Proc. Soft Computing and Signal Processing*, Singapore, Jan. 2019, pp. 135–143.