Computers in Biology and Medicine 150 (2022) 106084

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

RCMNet: A deep learning model assists CAR-T therapy for leukemia

Ruitao Zhang [a,b,1], Xueying Han [c,1], Zhengyang Lei [a,b,1], Chenyao Jiang [a,b], Ijaz Gul [a,b], Qiuyue Hu [a,b], Shiyao Zhai [a,b], Hong Liu [e], Lijin Lian [a,b], Ying Liu [e], Yongbing Zhang [f], Yuhan Dong [a,b], Can Yang Zhang [a,b], Tsz Kwan Lam [a,b], Yuxing Han [b], Dongmei Yu [d,***], Jin Zhou [c,**], Peiwu Qin [a,b,*]

[a] Institute of Biopharmaceutical and Health Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong 518055, China
[b] Precision Medicine and Public Health, Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen, Guangdong 518055, China
[c] The First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, China
[d] School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai, Shandong 264209, China
[e] Animal and Plant Inspection and Quarantine Technical Centre, Shenzhen Customs District, Shenzhen, Guangdong 518045, China
[f] Department of Computer Science, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China

A B S T R A C T

Acute leukemia is a type of blood cancer with a high mortality rate. Current therapeutic methods include bone marrow transplantation, supportive therapy, and chemotherapy. Although a satisfactory remission of the disease can be achieved, the risk of recurrence is still high. Therefore, novel treatments are demanding. Chimeric antigen receptor-T (CAR-T) therapy has emerged as a promising approach to treating and curing acute leukemia. To harness the therapeutic potential of CAR-T cell therapy for blood diseases, reliable cell morphological identification is crucial. Nevertheless, the identification of CAR-T cells is a big challenge posed by their phenotypic similarity with other blood cells. To address this substantial clinical challenge, herein we first construct a CAR-T dataset with 500 original microscopy images after staining. Following that, we create a novel integrated model called RCMNet (ResNet18 with Convolutional Block Attention Module and Multi-Head Self-Attention) that combines the convolutional neural network (CNN) and Transformer. The model shows 99.63% top-1 accuracy on the public dataset. Compared with previous reports, our model obtains satisfactory results for image classification. Although testing on the CAR-T cell dataset, a decent performance is observed, which is attributed to the limited size of the dataset. Transfer learning is adapted for RCMNet and a maximum of 83.36% accuracy is achieved, which is higher than that of other state-of-the-art models. This study evaluates the effectiveness of RCMNet on a big public dataset and translates it to a clinical dataset for diagnostic applications.

1. Introduction

Leukemia is a common hematopoietic malignant disease, which is hard to cure due to its malignant proliferation in the human body. According to global cancer statistics, 311,594 deaths and 474,519 new cases of leukemia have been reported worldwide in 2020 [1], and the number is rising gradually per year. The number of males reaches up to 269,503, which is higher than 205,016 for females and the ratio can be up to 1.31 [1]. Leukemia is categorized into two types, acute leukemia malignantly proliferating from the primitive white blood cells such as hematopoietic stem cells, and chronic leukemia malignantly proliferating from mature white blood cells. The worsening progress is faster and the death rate is higher for acute leukemia [2], leading to substantial therapeutic challenges. Generally, the methods to treat acute leukemia are bone marrow transplantation, supportive therapy, and chemotherapy. Although bone marrow transplantation is an acceptable approach to treating leukemia, some reports have demonstrated that once bone marrow transplantation is not enough, more transplants are required [3,4]. The research shows that bone marrow transplantation occurs with a high relapse incidence and low overall survival [5]. In addition, cure rates are related to the age stages, and the cure rate of childhood acute lymphoblastic leukemia can be up to over 90% [6]. Patients of some subtypes have lower survival rates as their immune cells cannot target the cancer cells. Novel therapeutic modalities to treat

* Corresponding author. Institute of Biopharmaceutical and Health Engineering, Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong 518055, China.
** Corresponding author.
*** Corresponding author.
E-mail address: pwqin@sz.tsinghua.edu.cn (P. Qin).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.compbiomed.2022.106084
Received 1 June 2022; Received in revised form 16 August 2022; Accepted 3 September 2022
Available online 11 September 2022

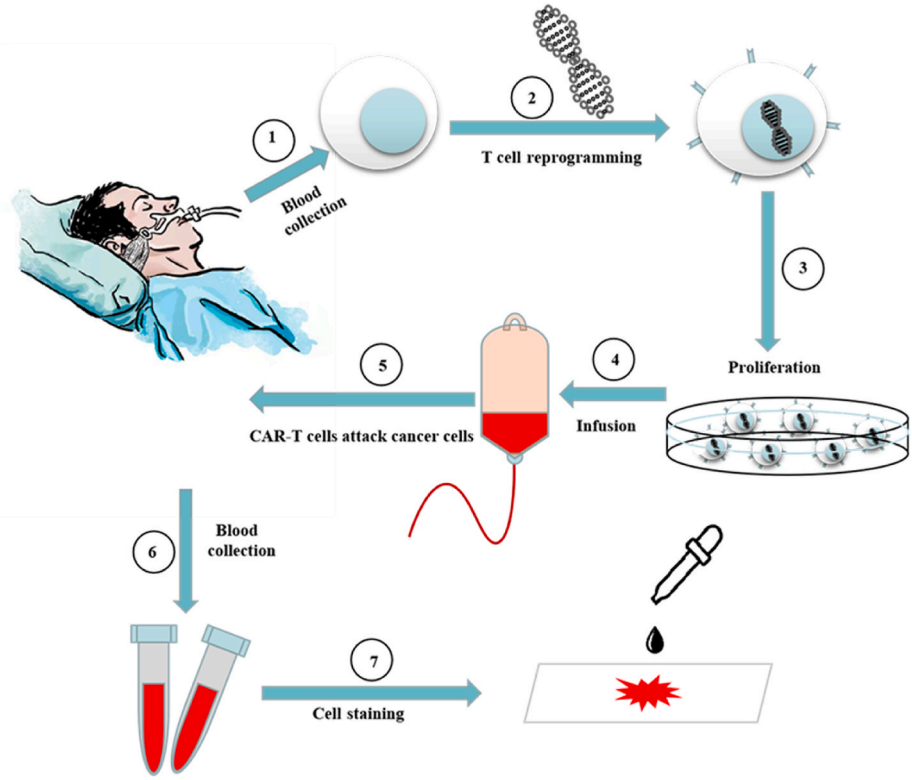

Fig. 1. The flow chart of CAR-T cell therapy and CAR-T staining for characterization.

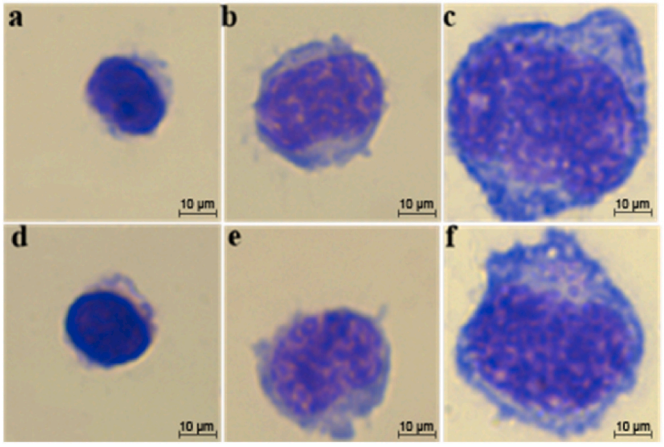

Fig. 2. Comparison between images of CAR-T cells and other cells with various shapes and sizes. (a–c) Images of CAR-T cells from 3 different patients. (d–f) Images of non-CART cells from normal blood samples. Scale bar is 10 μm.

leukemia patients are urgently needed.

CAR-T cell immunotherapy has shown the potential to address this challenge. A CAR-T cell is a type of artificial T cell, which is obtained by extracting the T cells from a leukemia patient and then modifying them by inserting a sequence of genes that can express receptor-recognizing tumor antigen and co-stimulatory molecules [8]. The CAR-T cells are then injected back into the patient. Fig. 1 illustrates the procedures of therapy and detection. For patients with leukemia, CAR-T cells can target the tumor cells specifically and consistently by proliferation to kill the abnormal cells compared with the drugs from chemotherapy that attack cells indistinguishably. Hence, CAR-T immunotherapy is deemed to be one of the most promising ways to cure leukemia and even other cancers. Clinical trials have shown that CAR-T can achieve a remission rate of 70–90% [7,8]. The complete elimination of blood cancer cells by CAR-T immunotherapy has been demonstrated, though the long-term evaluation is unknown [9,10].

Morphological identification of CAR-T cells is very important for blood diseases. Because of the individual differences, patients receiving CAR-T cell therapy may experience side reactions, which can be severely debilitating or fatal. It is extremely significant to instruct the clinical treatment by monitoring the development of immune cells in detail and real-time. A hematology analyzer can pinpoint the abnormal phenotypic cells rapidly. However, the automated analysis does not show better performance compared with manual identification, which is indispensable for the successful diagnosis of leukemia [11]. Although CAR-T immunotherapy is a promising treatment for leukemia, limited blood morphologists can identify the morphological features of the CAR-T cells precisely because of their heterogeneous nature in different patients and the similar phenotypes with other cells. Fig. 2 shows the examples of CAR-T cells and interfering blood cells with similar morphology, which is difficult to be distinguished by human eyes. Furthermore, the biochemical analysis or training of a novice blood morphologist who can recognize the CAR-T cells is time-consuming and expensive. The phenotype identification of CAR-T cells is crucial for disease prognosis. A survey has reported the partial characteristics of CAR-T cells, such as the large multinucleated forms [12]. However, there are no further reports on CAR-T cell characterization.

In recent years, deep learning (DL) has enticed substantial research attention by constructing models to assist doctors in diagnosing a variety of diseases. It has shown great success in medical cell image classification [13–15]. The traditional method of manual CAR-T cell classification based on morphological characteristics is a labor-intensive and time-consuming task [16]. DL methods for classifying CAR-T cells can help save time and make quick and accurate decisions. For image analysis, a reliable dataset is of paramount importance; the morphological observation is crucial for blood disease diagnosis and the recognition of CAR-T cells is significant for the prognosis of patients after receiving CAR-T therapy. Introducing a system for morphological

studies of CAR-T cells can be a good addition to the leukemia treatment modalities. This study comprises two parts: constructing a CAR-T cell dataset and proposing a novel DL model to distinguish CAR-T cells precisely. The contributions of our study are as follows:

- Construct the first CAR-T cell dataset from patients with leukemia with ethical approval.
- Design and implement a new multi-attention network for the CAR-T cell classification, which combines CNN with self-attention and applies it to CAR-T cell classification for the first time.

The organization of the paper is as follows: Section 2 shows relative work on cell classification. Section 3 illustrates the data collection and the method of building our model. Section 4 presents the result and discussion. Lastly, in section 5, the conclusion and future work are presented.

2. Relative work

2.1. CNN

Computer vision (CV) has shown increased applications in autonomous technology and advanced surveillance systems [17]. In recent years, medical image processing has enticed substantial research attention due to increasing shared medical resources, which can assist physicians in disease diagnosis and capture potential features that physicians rarely notice [13]. In contrast to doctors, CV can work continuously and efficiently, maintaining decent accuracy. The most classical method is machine learning (ML). Before ML classification, the pre-processing procedure extracts features with key information, such as principal component analysis (PCA) [18], independent component analysis (ICA) [19], and linear discriminant analysis (LDA) [20]. However, it is unable to capture all prominent features and thus loses some crucial information, resulting in lower accuracy. The classification methods can be divided into two categories: supervised learning and unsupervised learning. For the supervised learning techniques, support vector machine (SVM) [21] maps from the sample space to the feature space such that the distance between two classes is maximized. SVM has been demonstrated in the classification of blood cells [22] and protein-protein interaction prediction [23]. K-nearest neighbor (KNN) classifier [24] uses the classes of the k closest training samples to the new sample in the feature space to vote for its class and can be used to classify cervical cancer [25]. Classification and regression tree [26] is a decision tree method that uses the Gini index to automatically select classification features, employed as a feature selection method in cervical cell classification [27]. Among the unsupervised learning approaches, K-Means [28] is the most common approach. ML cannot obtain higher accuracy without sufficient information. Therefore, alternative methods are desired to overcome this challenge.

With the development of computer technology, novel hardware and software can provide and support larger memory and a higher hash rate for DL training. The appearance of DL resolves the limitations of ML. CNN is the major player in DL, which consists of many convolutional layers that contain convolutional kernels of various sizes, in an end-to-end learning schema. Via convolutional kernels, CNN can extract a mass of features randomly and assemble the key features for the downstream task. The application of CNN in medical image processing is appealing to both researchers and doctors. Y. Xie [29] proposed a dense convolutional neural network for stroke prediction via electrocardiogram (ECG). DL assists in the diagnosis of malignancies such as breast cancer, lung cancer, and malignant melanoma, accelerating diagnosis while lowering costs and workloads for clinical doctors [30]. For cell classification, outer phenotype and inner structure are two frequently-used input features. The reports about DL cell recognition include blood cells, cancer cells, and other cell types [27,31–33].

Different DL models have been adapted and created for cell recognition. A hybrid model that combines transfer learning with generative adversarial networks (GANs) increases the accuracy of a small dataset with staining-free cancer cells. They collect the optical path delay maps from low-coherence off-axis holography as input and pretrain the GANs with sperm cells before training their dataset. The accuracy of the model is 99%, which outperforms the single GANs or MobileNet with transfer learning [31]. Another report achieves an accuracy of 99.54% by combining convolutional deep neural network and SVM to classify the sickle cells and normal blood cells with transfer learning and data enhancement [32]. A deeply supervised residual network can classify human epithelial-2 cells with an accuracy of 99.98% [33]. For the cellular inner structure, a study shows that DL can classify the *Cercopithecus aethiops* monkey kidney cells based on the microtubule networks, and it shows better results than the human expert [34]. Another research reveals that CNN can distinguish the normal and cancer cells in the breast by recognizing discrepancies in the actin cytoskeleton structures that can serve as a supererogatory diagnostic marker, which outperforms the human experts [13].

2.2. CNN and transformer

CNN owns powerful inductive biases, such as local correlation and weight sharing, which improve the accuracy and effectiveness. However, it limits the performance upper bound of the model as well. Although deeper CNN can weaken the effects of the limited receptive field and long-range dependence, more complex and larger CNN is needed, which increases the difficulty of training. The Transformer performing well in global correlation is extensively used in natural language processing (NLP) and has achieved great success. However, Transformer training is time-consuming and requires an extremely large dataset and high computational memory. Significant efforts have been made to combine CNN and the Transformer to balance each other to get better results. Vision Transformer (ViT) [35] crops several patches from the image and reshapes the patches following the word embedding as input. The author maintains the Transformer structure and changes the input that imitates the word embedding. Bello et al. propose a hybrid model with CNN and Transformer, named AA-ResNet, with 77.7% accuracy on ImageNet classification [36]. Nevertheless, a comprehensive Transformer is complex and less flexible to be transferred to image processing from text processing. Bottleneck Transformer (Bot) block [37] utilizes the core self-attention from Transformer to replace the middle convolutional layer of the last blocks from ResNet50. Although the model structure doesn't change too much, it achieves higher accuracy compared with the traditional CNN.

2.3. CAR-T dataset

To the best of our knowledge, a CAR-T cell dataset is unavailable so far. The establishment of the CAR-T database requires the collaboration of experienced morphologists with special expertise. In this work, the CAR-T cell dataset is constructed for the first time, which can be used as the baseline and reference for the later CAR-T cell dataset construction and research.

3. Methodology

3.1. Dataset

We train and test our model and a few popular models for cell recognition on two datasets with different cell types. The first one is a common dataset for microscopic peripheral blood cells and the other one is our dataset for CAR-T cells obtained with ethical approval and patient notification letter.

3.1.1. Peripheral Blood Cells (PBC) dataset

Acevedo et al. published the PBC dataset in 2020 [38], which is one

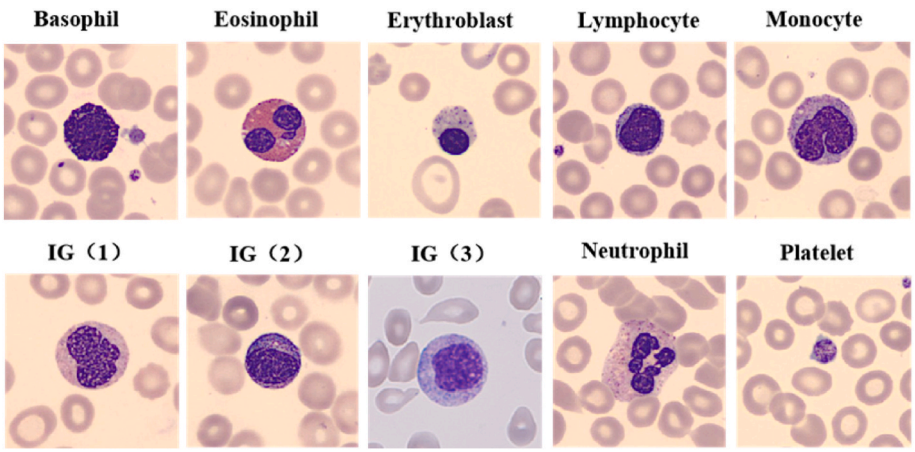

Fig. 3. Representative images from 8 classes of PBC dataset showing Basophils, Eosinophils, Erythroblasts, Lymphocytes, Monocytes, IGs, Neutrophils, and Platelets. IG1, IG2 and IG3 represent the metamyelocytes, myelocytes, and promyelocytes respectively.

Table 1
The cell type, number, percentage, and the number of images used for the train and test set for the PBC dataset.

Cell type	Number	Percentage	Train Set	Test Set
Neutrophils	3329	19.48%	971	243
Eosinophils	3117	18.24%	971	243
Immature granulocytes (Metamyelocytes, Myelocytes and Promyelocytes)	2895	16.94%	971	243
Platelets (Thrombocytes)	2348	13.74%	971	243
Erythroblasts	1551	9.07%	971	243
Monocytes	1420	8.31%	971	243
Basophils	1218	7.13%	971	243
Lymphocytes	1214	7.10%	971	243
Total	17,092	100%	7768	1458

Table 2
CAR-T cell dataset.

The Type of Cell	Number	Percentage	Train Set	Test Set
CAR-T cell	250	50%	200	50
Normal cell	250	50%	200	50
Total	500	100%	400	100

Table 3
CAR-T cell dataset after data augmentation.

The Type of Cell	Number	Percentage	Train Set	Test Set
CAR-T cell	1500	50%	1200	300
Normal cell	1500	50%	1200	300
Total	3000	100%	2400	600

of the largest and most complete available datasets about blood cells. The dataset is collected by the core laboratory at the hospital clinic of Barcelona with the analyzer Cella Vision DM96, where the May Grünwald-Giemsa stain [39] is used to stain the cells in the autostainer Sysmex SP1000i. The PBC dataset contains eight cell types, including neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (IGs), erythroblasts, and platelets (Fig. 3). There are three subtypes of IG consisting of promyelocytes, myelocytes, and metamyelocytes, which makes it difficult to differentiate the immature granulocytes from other cells. The total number of images is 17,092. The details of the dataset are shown in Table 1. The cells are labeled by professional clinical pathologists and the image size is 360 × 363 pixels. Each image contains one cell. For the PBC dataset, to minimize the cell type imbalance that may influence the learning effectiveness of the model, the number of lymphocytes is regarded as standard and other cell types are truncated to the same number and all selected images are chosen randomly from the corresponding categories. Table 1 presents the descriptive details of all types of cells.

3.1.2. CAR-T cell dataset

Six patients with acute lymphoblastic leukemia receive anti-CD19 CAR-T therapy and clinical biopsy from the first affiliated hospital of Harbin Medical University. The blood samples are collected after the CAR-T therapy within several days or weeks. All patients have signed an informed agreement letter to provide blood samples for this study. The protocol gains ethical approval from the hospital ethical committee. All blood samples, which are gathered from the patients receiving CAR-T cell therapy, are stained by May Grünwald-Giemsa (*39*) and the CAR-T cells have been confirmed by immunostaining [12]. Cells are transferred to glass bottom dish (150680 Thermo Scientific, USA) and fixed with 4% paraformaldehyde at 37 °C for 15 min. Blocking buffer (P0102 Beyotime Biotechnology, China) is used to treat cells for 1 h at room temperature, then incubated with primary antibodies (FM3-S93 Acro-Biosystems, USA) for 1 h at room temperature and with secondary antibodies (A0562 Beyotime Biotechnology) for 1 h at room temperature. Nuclei are stained with DAPI (C1005 Beyotime Biotechnology) in PBS for 5 min. To collect CAR-T images, we use a wide-field microscopy with a 100x oil immersion lens (Leica, DM500). The size of each image is 384 × 384 pixels with only one cell. Because of the complexity and scarcity of blood, we label the CAR-T cells with the help of a professional blood morphologist. For our dataset, there are two categories of cells, including the CAR-T and the other cells. We collect 250 pictures per class. The assignment ratio is 8:2 for each class, which means that there are 200 images for the training set and 50 samples for the test set. The dataset description is shown in Table 2.

3.2. Data augmentation

A limited number of images may lead to the model overfitting. Dataset expansion is an efficient methodology to decrease overfitting. A similar strategy has been reported earlier [40]. An artificial neural network model is used to classify white blood cells, where data augmentation enhances the model performance and can reduce overfitting [41]. We utilize data augmentation to increase the size of the limited dataset with rotations and flips including three rotations: 90°, 180°, and 270°, and two flips: vertical and horizontal flips. The final size of the CAR-T dataset expands from 500 to 3000 images with approximately 1500 images per class to balance the dataset (Table 3).

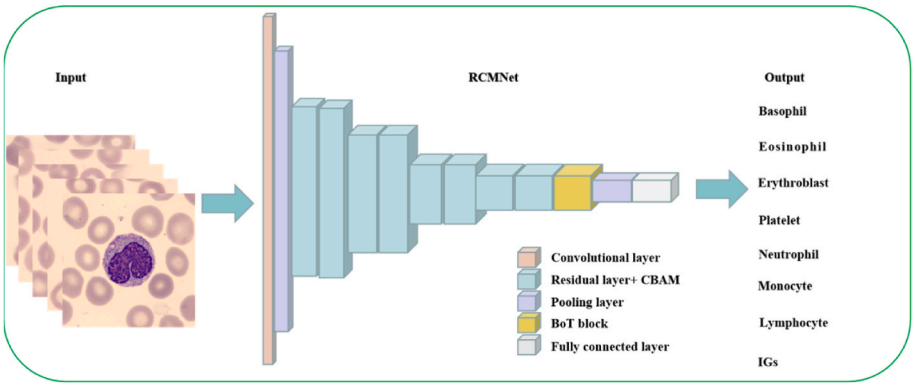

Fig. 4. RCMNet schematic. The microscopy images (360 × 363 or 384 × 384) are utilized as input. CBAM and BoT block are inserted into ResNet18. The output is the result of cell classification.

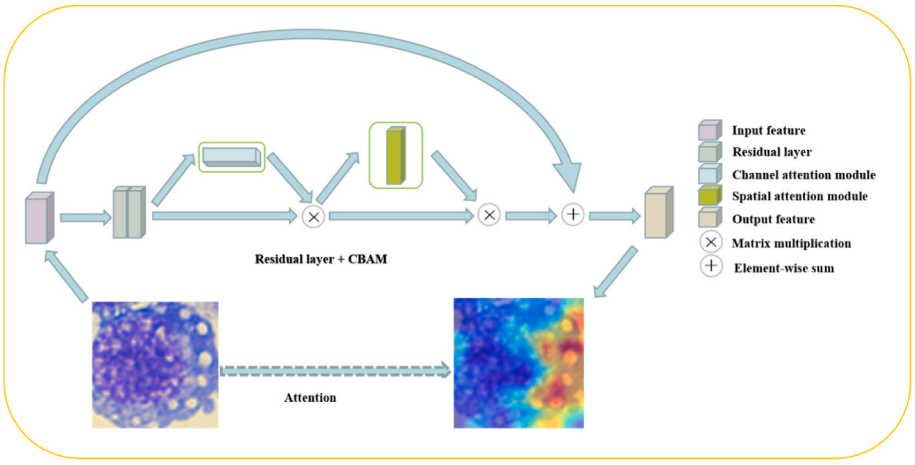

Fig. 5. The combination between residual block and CBAM.

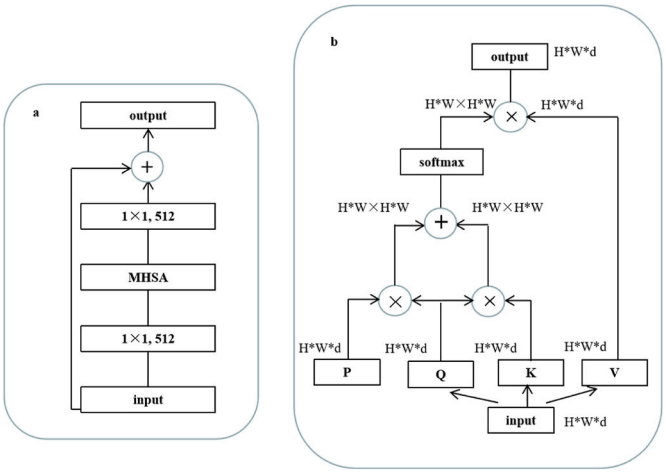

Fig. 6. Diagram illustrating the BoT block. The structure of BoT block (a) and self-attention (b). Q, K, V, and P represent query, key, value, and position embedding separately. (×) represents matrix multiplication and (+) represents element-wise sum.

3.3. Model

The schematic of the proposed model is illustrated in Fig. 4. PBC and CAR-T datasets are used as input separately. The model output is the cell recognition based on the features extracted after a series of convolutional operations. ResNet18 serves as the backbone with two inserted blocks consisting of Convolutional Block Attention Module (CBAM) and Multi-Head Self-Attention (MHSA), which are two key attention blocks in our model. Details of the model are described in the following section.

3.3.1. Residual neural network (ResNet)

ResNet is a common and classical neural network in computer vision [42]. Researchers usually use ResNet as the backbone and modify it for different purposes. There are plenty of ResNet variants, such as ResNeXt [43], DenseNet [44], and MobileNet [45]. A unique property of ResNet is the shortcut connection to maintain the original features while addressing the challenge of exploding gradient and vanishing gradient. ResNet is adopted as the backbone of our model, composed of a series of residual blocks. The functions depicting the learning process are shown as below:

$$x_{l+1} = Relu(x_l + f(x_l, w_l)) \tag{1}$$

where x_l and x_{l+1} represent the input for the l_{th} layer and output for the $(l+1)_{th}$ layer, $Relu(\cdot)$ is the rectified linear unit function, $f(\cdot)$ represents the residual mapping function, w_l is the parameters for the l_{th} layer.

Based on Equation (1), the cumulative operations (Equation (2)) up to the L_{th} layer can be represented as the following:

$$x_L = x_l + \sum_{i=l}^{L-1} f(x_i, w_i) \tag{2}$$

where x_L represents the output for the L_{th} layer, which compresses the

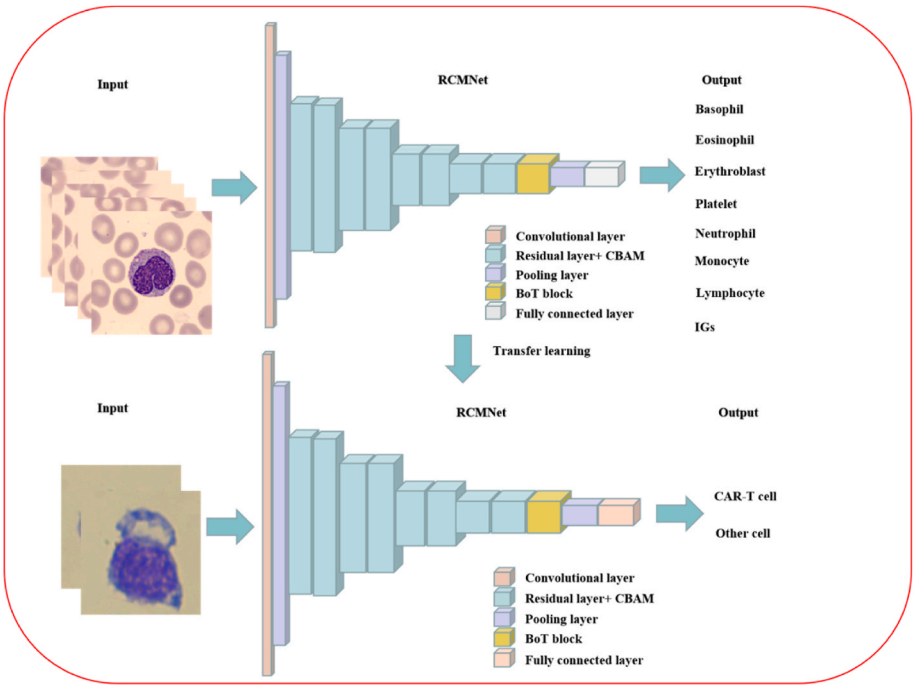

Fig. 7. The schematic of transfer learning.

Table 4
Comparison among various models: ResNet18, ResNet18-M, ResNet18-C and RCMNet. ResNet18 is the backbone. Resnet18-C and ResNet18-M are the variants from ResNet18. RCMNet is our proposed model.

	ResNet18	ResNet18-C	ResNet18-M	RCMNet
Conv1	7 × 7 conv,64, Stride2 3 × 3 Max Pooing, Stride2			
Layer1	[3 × 3 conv, 64 3 × 3 conv, 64] × 2	[3 × 3 conv, 64 3 × 3 conv, 64 CBAM, 64] × 2	[3 × 3 conv, 64 3 × 3 conv, 64] × 2	[3 × 3 conv, 64 3 × 3 conv, 64 CBAM, 64] × 2
Layer2	[3 × 3 conv, 128 3 × 3 conv, 128] × 2	[3 × 3 conv, 128 3 × 3 conv, 128 CBAM, 128] × 2	[3 × 3 conv, 128 3 × 3 conv, 128] × 2	[3 × 3 conv, 128 3 × 3 conv, 128 CBAM, 128] × 2
Layer3	[3 × 3 conv, 256 3 × 3 conv, 256] × 2	[3 × 3 conv, 256 3 × 3 conv, 256 CBAM, 256] × 2	[3 × 3 conv,256 3 × 3 conv, 256] × 2	[3 × 3 conv, 256 3 × 3 conv, 256 CBAM, 256] × 2
Layer4	[3 × 3 conv, 512 3 × 3 conv, 512] × 2	[3 × 3 conv, 512 3 × 3 conv, 512 CBAM, 512] × 2	[3 × 3 conv,512 3 × 3 conv, 512] × 2	[3 × 3 conv, 512 3 × 3 conv, 512 CBAM, 512] × 2
Layer5	– Average Pooling, 2-FC	–	BoT Block	BoT Block

middle layers by summing the shallow residual block.

3.3.2. CBAM

CBAM is an attention mechanism proposed by Woo et al. [46], derived from Squeeze-and-Excitation Networks (SENet) [47]. Compared with the common model modifications, such as increasing the depth and width, CBAM is a lightweight module, which doesn't consume too much computational memory to achieve higher accuracy. CBAM can be easily inserted into any layer because of the module's characteristic. We incorporate CBAM into each residual block of our model. Fig. 5 shows the combination of residual block and CBAM module, which is introduced after two convolutional layers. There are two attention modules including the channel attention module and the spatial attention module, where the channel attention module is ahead of spatial attention. Channel attention can be regarded as the feature extractor for each channel. According to fully connection calculating the assigned weight of each channel, the significant features, such as texture, outline, etc., will be integrated and compressed to the size of 1 × 1 × n (n represents the number of classifications) after average pooling and max pooling. The functions for channel attention (Equations (3)–(5)) are shown as below:

$$\varphi_a = F_c(Relu(F_c(avg(m)))) \tag{3}$$

$$\varphi_m = F_c(Relu(F_c(max(m)))) \tag{4}$$

$$\varphi = \sigma(\varphi_a + \varphi_m) \tag{5}$$

where m is input, $avg(\cdot)$ is average pooling based on width, $max(\cdot)$ is max pooling based on height, φ_a is the output after MLP for average

Table 5
Comparison among ResNet18, ResNet18-M, ResNet18-C, VGG19, AlexNet, and previously published works and RCMNet under the same setting in Top-1 accuracy and Top-5 accuracy. All images with the resolution of 360 × 363 are trained for 30 epochs. NA is not available.

Model	Top-1 acc.	Top-5 acc.
ResNet18	99.51	100
ResNet18-M	99.48	100
ResNet18-C	99.58	100
RCMNet (Ours)	99.63(+0.12)	100
VGG19	99.25	100
AlexNet	99.18	100
Acevedo et al. [53]	96.20	NA
Ucar [54]	97.94	NA
Long et al. [55]	99.30	NA

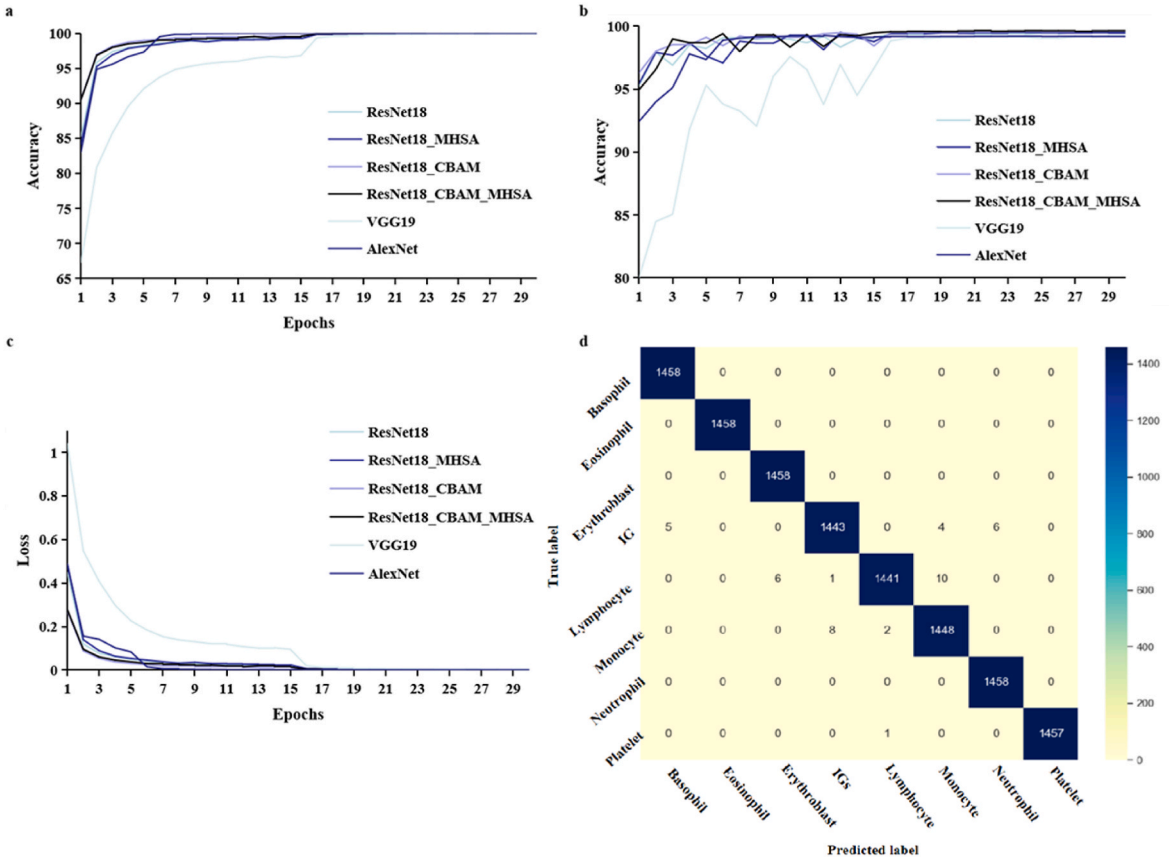

Fig. 8. Training (a) and testing (b) accuracy, training loss (c) and confusion matrix (d) on the PBC dataset. The values on the diagonal of the confusion matrix indicate the number of correct predictions and the off-diagonal values are the numbers of samples with incorrect predictions.

Table 6
Comparison of ResNet18, ResNet18-M, ResNet18-C, VGG19, AlexNet, and RCMNet under the same settings in Top-1 accuracy. All images with the resolution of 384 × 384 are trained for 20 epochs.

Model	Top-1 acc.
ResNet18	81.24
ResNet18-M	78.22
ResNet18-C	82.63
RCMNet	80.01
VGG19	79.11
AlexNet	78.46

pooling feature, φ_m is the output after MLP for max pooling, $F_c(\cdot)$ is fully connection, $\sigma(\cdot)$ is the sigmoid function, φ is the output of channel attention.

For spatial attention, it calculates the inner relationship between pixels to decide the focus on the image. Using convolution with a 7 × 7 convolutional kernel, the features after max pooling and average pooling will be compressed to the size of W × H × 1 (W represents the width, H is the height) to determine the assignment of spatial weight. The function for special attention (Equation (6)) is performed as follows:

$$\varepsilon = \sigma(conv(conc(avg(m), max(m)))) \tag{6}$$

where ε represents the output of special attention, $conv(\cdot)$ is the 7 × 7 convolutional kernel, $conc(\cdot)$ stands for the concatenation between average pooling and max pooling.

3.3.3. MHSA

The Bottleneck Transformer Network (BoTNet) is proposed by Srinivas et al. [37]. The middle convolutional layer of the last bottleneck block can be replaced by MHSA in BoT block [48], which is commonly used in NLP. The structure of the BoT block is shown in Fig. 6a [48]. In contrast to MHSA for CNN, Transformer with MHSA utilizes layer normalization, single non-linearity, Adam optimizer, and output projection, while the BoT block uses batch normalization, three non-linearities, SGD optimizer [37], and there is no output projection in the BoT block. 2D relative position encoding [49] is applied to the MHSA of the BoT block to pinpoint the position for each pixel. Fig. 6b shows the details of self-attention. Because Transformer can consume a large amount of computational memory, insertion of the BoT block into the last layer of CNN decreases the memory consumption. Inspired by the BoTNet design, we insert the BoT block with 4-head MHSA after the final layer of the ResNet18. The structure is shown in Fig. 6.

3.4. Transfer learning

Because the doctors have no time to label large number of medical images, which is time-consuming and there are not enough images in most cases, limited medical data are acquired. A small dataset usually doesn't offer efficient information for the training model leading to overfitting. Transfer learning alleviates the impact of small data size, which is a promising method to increase the model's accuracy. Transfer learning transfers the weights learned in previous training with a large public dataset to the target domain, accelerating the learning efficiency without training from scratch. The precondition for transfer learning is similar to data distributions such that prior knowledge can be transferred to the novel model. Raghu et al. [50] have verified that the transfer learning with ResNet50 shows better results for medical image processing on a more complex model. In this study, transfer learning is adopted because there are limited images for the CAR-T dataset. For

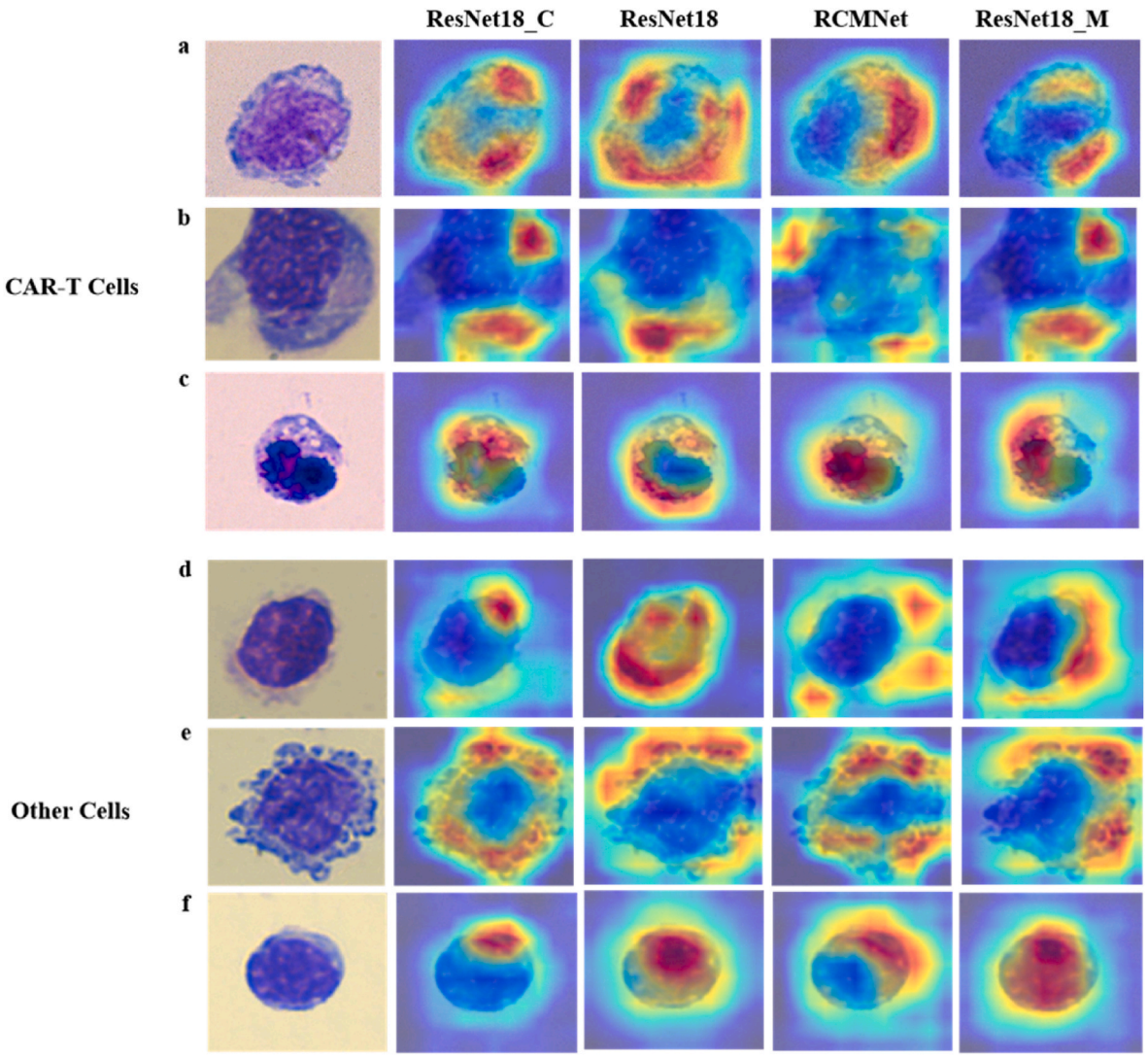

Fig. 9. Grad-CAM output from ResNet18, ResNet18-M, ResNet18-C, RCMNet without transfer learning on the CAR-T cell dataset. (a–c) Grad-CAM output of CAR-T cells. (d–f) Grad-CAM output of other blood cells.

Table 7
Comparison of ResNet18, ResNet18-M, ResNet18-C, VGG19, AlexNet, and RCMNet under the same settings in Top-1 accuracy after transfer learning. All images with the resolution of 360 × 363 are trained for 20 epochs.

Model	Top-1 acc.
ResNet18	82.18
ResNet18-M	82.13
ResNet18-C	83.01
RCMNet	**83.36**
VGG19	79.69
AlexNet	79.03

RCMNet, the pre-train model is trained on the PBC dataset. Due to the similarity between the PBC dataset and the CAR-T dataset, all weights learned with the PBC dataset except a fully connected layer are transferred and re-trained on the CAR-T dataset. The parameters from the fully connected layer are initialized randomly. Finally, the pre-train model is re-trained on the CAR-T dataset by freezing weights from all layers without the fully connected layer and fine-tuning the weight of the fully connected layer. The schematic is shown in Fig. 7.

3.5. Gradient-weighted class activation mapping (Grad-CAM)

Class activation mapping (CAM) [51] visualizes deep learning features, generates the thermodynamic region, and illustrates the importance of relevance for decision-making in deep learning. Grad-CAM [52] doesn't change the structure or retrain the model and calculates the gradient of A^i_{mn} according to backpropagation. We obtain the average gradient value by global average pooling for each feature map, which is the α^c_i. α^c_i times the corresponding feature map before *Relu* to get the final Grad-CAM. The functions for Grad-CAM (Equations (7) and (8)) are shown below:

$$\alpha^c_i = \frac{1}{Z}\sum_{m=1}^{w}\sum_{n=1}^{h}\frac{\partial S_c}{\partial A^i_{mn}} \tag{7}$$

$$L^c_G = Relu\left(\sum_i \alpha^c_i A^i\right) \tag{8}$$

α^c_i represents the sensitivity of the i_{th} channel of feature map in the final convolutional layer; i represents the index of the channel; c represents the index of classification; A represents the feature map from the final convolutional layer; S represents the probability vector; Z represents the size of feature ma; m *and* n represent the height and width; *Relu*

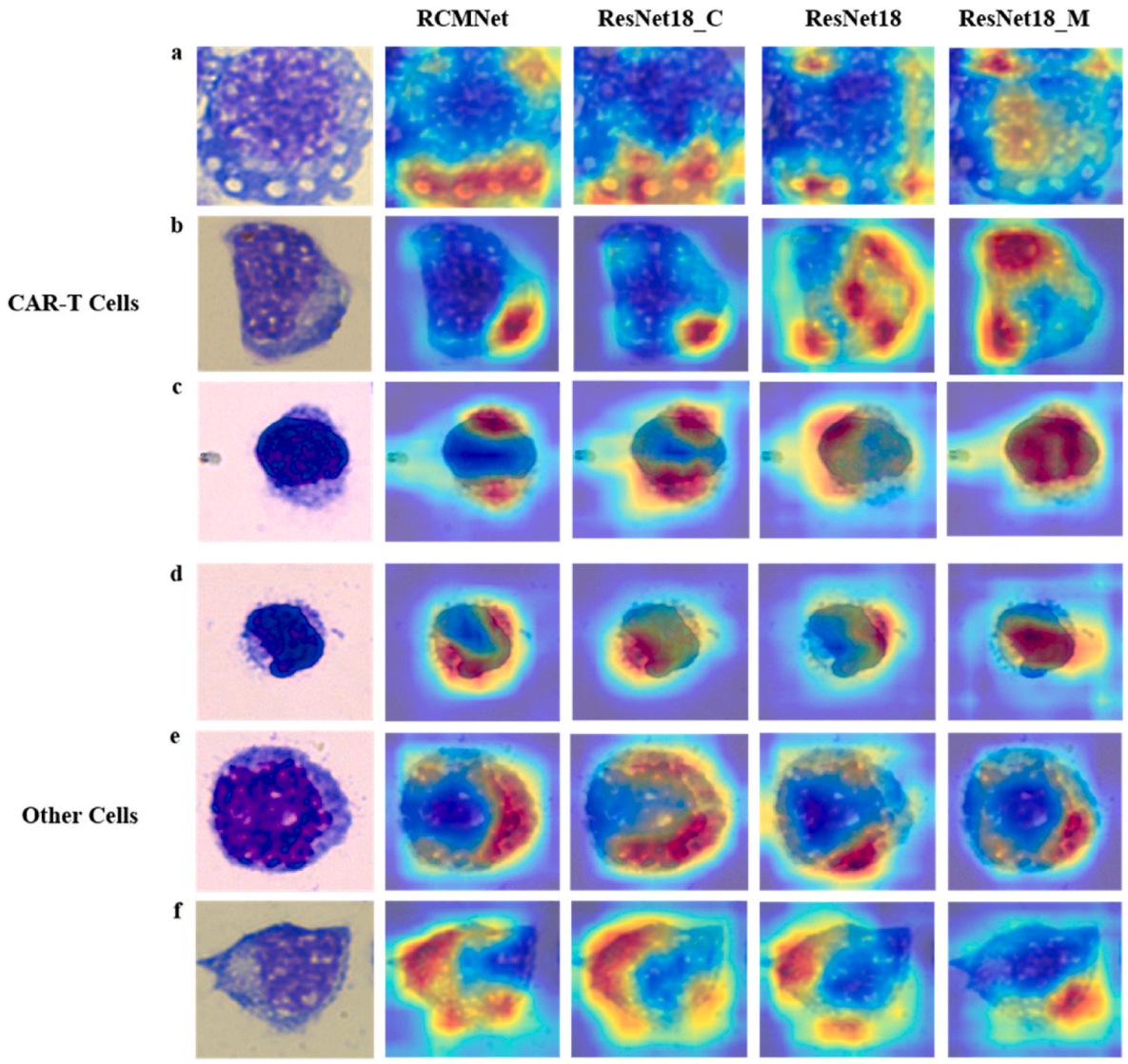

Fig. 10. Grad-CAM output from ResNet18, ResNet18-M, ResNet18-C, RCMNet with transfer learning on CAR-T cell dataset. (a–c) Grad-CAM output of CAR-T cells. (d–f) Grad-CAM output of other cells.

represents the activation function; L_G^c represents the Grad-CAM.

4. Results

To evaluate the performance of our proposed model, we test the model with two datasets: the PBC dataset and the novel CAR-T dataset. PBC dataset is the benchmark to evaluate the capability of our model with available models and we create an original CAR-T dataset to demonstrate the applicability of our model for the clinical challenges. We compare the performance of our model with other common models (e.g., ResNet, AlexNet, and VGG19). Meanwhile, various ablation experiments are designed to demonstrate the effectiveness of our method on the PBC dataset. Top-1 and Top-5 accuracy levels are reported. All training and testing tasks are operated on the NVIDIA GeForce RTX 2080 SUPER.

4.1. Image classification for the PBC dataset

4.1.1. Ablation study

To evaluate the effectiveness of the final architecture of the model, we conduct an ablation experiment. ResNet18 is the backbone and all modifications are based on ResNet18. All the details of the model architecture are explained in Table 4. The designed ResNet18 variant is the model without CBAM and MHSA (ResNet18), the model with CBAM but without MHSA (ResNet18-M), the model with MHSA but without CBAM (ResNet18-C), and the model with MHSA and CBAM (RCMNet).

We test different models on the public PBC dataset, and all hyper-parameters are the same with 30 epochs across the whole training schedule (Table 5). RCMNet has a great improvement compared with the other three models, which indicates that the combination of CBAM and MHSA is effective for image classification for the PBC dataset. 2-FC represents the full connection with 2 types of output.

4.1.2. Model performance comparison with the PBC dataset

After evaluating the effectiveness of our proposed model by an ablation experiment, additional comparison analyses are carried out. Table 5 shows the test accuracy for all models. The accuracy of AlexNet is the lowest; our model has the highest accuracy. RCMNet improves on top of AlexNet by 0.45% in the regular setting. Here, the accuracy of ResNet18_M is lower than that of ResNet18 and ResNet18_C, which suggests that the CBAM can assist ResNet18 to aggregate features with effective information, achieving better performance after inserting CBAM into the BoT block. Training and testing accuracy, training loss, and confusion matrix of RCMNet classifier after 30 epochs are shown in

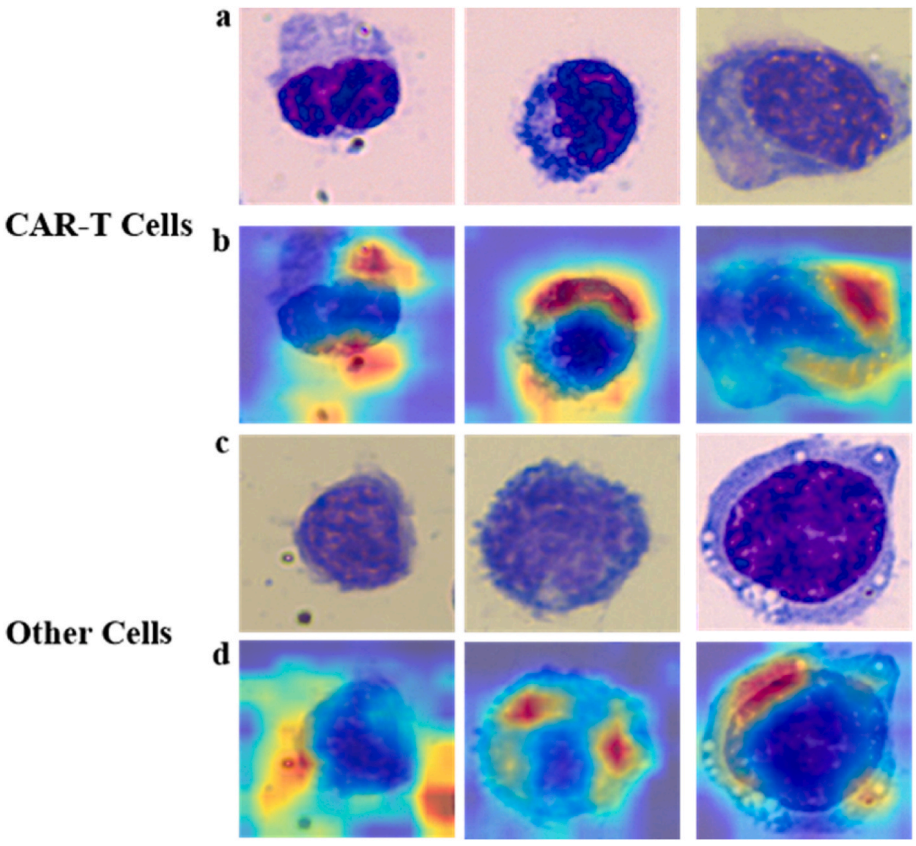

Fig. 11. The failed classification cases from RCMNet after transfer learning on the CAR-T cell dataset. (a–b) Images and Grad-CAM output of CAR-T cells. (d–e) Images and Grad-CAM output of other cells.

Fig. 8. All models converge after 17 epochs (Fig. 8a–c). In Fig. 8d, we find that the sub-dataset of IGs shows the most significant confusing score for our model. We speculate that three cell types of promyelocytes, myelocytes, and metamyelocytes belonging to IGs have discernible different phenotypes. Meanwhile, our model achieves higher accuracy with the PCB dataset compared with the previous work.

4.2. Image classification for CAR-T dataset

4.2.1. Comparison of the CAR-T dataset

The ablation experiments show promising results on the common dataset with great improvement. Next, we study the performance of RCMNet on the CAR-T dataset. We compare the RCMNet model with common models and ResNet18 variants and Top-1 accuracy after 20 epochs are presented in Table 6. Our model shows at least comparable or relatively higher accuracy than most of the models. However, ResNet18 and Resnet18-C outperform RCMNet in CAR-T dataset. Fig. 9 displays the visualization comparison for different models, where RCMNet and ResNet18-M show worse performance, consistent with the calculated results. The focus attention for RCMNet and ResNet18-M is more dispersive. We speculate that because MHSA cannot extract features with enough information for a small dataset causing self-attention incapable of integrating global information comprehensively. Another paper reports that transformer-based models perform worse than the CNN-based models on small datasets but outperform CNN-based models on larger datasets (*35*). In addition, RCMNet can lose some features in a small dataset, especially for the tiny features during the progress of convolution while they are the keys for classification.

4.2.2. Enhanced performance of RCMNet via transfer learning

Direct RCMNet model training seemly shows that the accuracy is not ideal due to the limited sample size. The disadvantages and limitations of a small dataset have been discussed extensively since the model cannot get enough training and generalization causing terrible results in predicting unknown cases. Transfer learning can increase the generalization performance between two similar datasets. Table 7 displays the results for each model after transfer learning. In this study, transfer learning is applied to RCMNet and 83.36% accuracy is achieved. Compared with RCMNet without transfer learning, the performance improvement for RCMNet with transfer learning is 3.35%, which is a satisfactory improvement. This result indicates that training RCMNet on CAR-T dataset without transfer learning cannot fully utilize the power of RCMNet in cell recognition. Fig. 10 shows the output of Grad-CAM. RCMNet with transfer learning is easier to focus on the key point with fewer redundant parts. Some failed cases are shown in Fig. 11 and the dispersive cases are remaining, which is a challenge for further study. In addition, the capacity of transfer learning for VGG19, ResNet18, and other models is very limited and the increase is not notable. This result demonstrates the potential of transfer learning in the application of self-attention for medical image processing when a large deep learning model is deployed and a small data size is available.

5. Conclusion

Due to the end-to-end inference and efficient feature extraction of deep learning, CAR-T cell recognition is an appropriate and meaningful target due to its clinical value. In this work, we make two unique contributions to the field of CAR-T cell recognition, which still relies on human labor to differentiate the presence of CAR-T cells after treatment [1]. We successfully construct a CAR-T dataset containing 500 CAR-T cell images, which have been labeled by an experienced blood morphologist [2]. The second one is that we construct a model named RCMNet with two attention mechanisms for classifying the CAR-T cell dataset. To the best of our knowledge, we create the first dataset that can

be used to develop a classification algorithm for CAR-T cells. Our proposed model RCMNet shows the potential for assisting doctors in identifying CAR-T cells and makes a crucial contribution to clinical decision-making in the selection of the treatment for acute leukemia. RCMNet is a hybrid model consisting of CNN and self-attention. The biggest benefit is the integration of the local and global information extracted from the images. On the PBC dataset, our model achieves 99.63% top-1 accuracy, outperforming the previously published works. Our model performs well in identifying CAR-T cells with minor morphological differences from non-CAR-T cells. Therefore, DL has the potential to aid in diagnosis of leukemia based on morphological distinction [56]. The model with local and global attention mechanism has the potential to aid in the diagnosis of various types of malignancies, such as lung cancer and breast cancer, since CAR-T cells are challenging to be identified due to their similar phenotypes with interfering cells. However, the model is sample size dependent and a large amount of data is preferred. Although the results of our model on the small CAR-T dataset cannot fully exploit its advantage, the accuracy on the single ResNet18-C still achieves 82.63%, and transfer learning is a valid method to increase the result to 83.36% accuracy on the RCMNet. This is the first report on CAR-T cell recognition, which is more challenging than regular blood cell classification. The high demand for CAR-T cell therapy urgently requires a platform and method that can facilitate CAR-T cell recognition and clinical decision-making. In the future, the CAR-T dataset can be expanded to include more expert-labeled images, test the model performance on larger datasets like ImageNet [57] and consider replacing the backbone with other architectures, such as inception-v4 [58]. The complex model architecture and transfer learning on ImageNet can be adapted to improve the model performance in CAR-T cell diagnosis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by grants from National Natural Science Foundation of China (31970752), Science, Technology, Innovation Commission of Shenzhen Municipality (JCYJ20190809180003689, JSGG20200225150707332, WDZC20200820173710001, and JSGG20191129110812708), and Shenzhen Bay Laboratory Open Funding (SZBL2020090501004).

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, L. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (2021) 209–249.
[2] M.F. Greaves, Aetiology of Acute Leukaemia, Lancet, 1997.
[3] A.Y. Leung, E. Tse, Y.Y. Hwang, T.S.Y. Chan, H. Gill, C.S. Chim, A.K.W. Lie, Y. L. Kwong, Primary treatment of leukemia relapses after allogeneic hematopoietic stem cell transplantation with reduced-intensity conditioning second transplantation from the original donor, Am. J. Hematol. 88 (2013) 485–491.
[4] M.M. Al Malki, L. Aldoss, T. Stiller, R. Nakamura, D. Snyder, S.J. Forman, V. Pullarkat, Outcome of second allogeneic hematopoietic cell transplantation in patients with acute lymphoblastic leukemia, Clin. Lymphoma, Myeloma & Leukemia 16 (2016) 519–522.
[5] A. Nagler, M. Labopin, B. Dholaria, J. Finke, A. Brecht, U. Schanz, R. Niittyvuopio, A. Neubauer, M. Bornhäuser, S. Santarone, D. Beelen, A. Shimoni, W. Rösler, S. Giebel, B. Savani, M. Mohty, Second allogeneic stem cell transplantation in patients with acute lymphoblastic leukaemia: a study on behalf of the Acute Leukaemia Working Party of the European Society for Blood and Marrow Transplantation, Br. J. Haematol. 186 (2019) 767–776.
[6] C.-H. Pui, Precision medicine in acute lymphoblastic leukemia, Front. Med. 14 (2020) 689–700.
[7] M. Sadelain, R. Brentjens, M. Davila, I. Riviere, S.J.C.R. Giralt, Abstract CT102: Efficiancy Toxicity Manage. Cell Therapy Acute Lymphoblastic. Leukemia 74 (2014). CT102-CT102.
[8] K.C. Pehlivan, B.B. Duncan, D.W. Lee, CAR-T cell therapy for acute lymphoblastic leukemia: transforming the treatment of relapsed and refractory disease, Curr. Hematol. Malig. Rep. 13 (2018) 396–406.
[9] D.W. Lee, J.N. Kochenderfer, M. Stetler-Stevenson, K.Y. Cui, S.A. Delbrook, T. J. Feldman, R. Fry, M. Orentas, N.N. Sabatino, S.M. Shah, Steinberg, D. Stroncek, N. Tschernia, C. Yuan, H. Zhang, L. Zhang, S.A. Rosenberg, A.S. Wayne, C. L. Mackall, T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial, Lancet 385 (2015) 517–528.
[10] R.A. Gardner, O. Finney, C. Annesley, H. Brakke, C. Summers, K. Leger, M. Bleakley, C. Brown, S. Mgebroff, K. Spratt, Intent-to-treat leukemia remission by CD19 CAR T cells of defined formulation and dose in children and young adults, Blood 129 (2017) 3322–3331.
[11] X.M. Chen, Meaning analysis of blood cell morphological observation to the common blood disease diagnosis, Chin Modern Med. 36 (2015) 100–102.
[12] X. Han, C. Wang, J. Zhou, Chimeric antigen receptor T (CAR-T) cells present with reactive and pleomorphic morphology in bone marrow, Am. J. Hematol. 94 (2019) 1297–1298.
[13] R.W. Oei, G. Hou, F. Liu, J. Zhong, J. Zhang, Z. An, L. Xu, Y. Yang, Convolutional neural network for cell classification using microscope images of intracellular actin networks, PLoS One 14 (2019), e0213626.
[14] A.I. Shahin, Y. Guo, K.M. Amin, A.A. Sharawi, White blood cells identification system based on convolutional deep neural learning networks, Comput. Methods Progr. Biomed. 168 (2019) 69–80.
[15] M.A.R. Ridoy, M.R. Islam, ICCIT), 2020 paper presented at the 2020 23rd International Conference on Computer and Information Technology.
[16] Y. Hu, J. Huang, The chimeric antigen receptor detection toolkit, Front. Immunol. 11 (2020) 1770.
[17] M. Wang, W.J.a. Deng, Deep Face Recognition: A Survey, 2018.
[18] S. Nazlibilek, D. Karacor, White blood cells classifications by SURF image matching, PCA and dendrogram, Biomed. Res. 26 (2015) 633–640.
[19] M.O. Adebiyi, A.A. Adebiyi, O. Okesola, M.O. Arowolo, ICA learning approach for predicting RNA-seq data using KNN and decision tree classifiers, Int. J. Adv. Sci. Technol. 29 (2020) 12273–12282.
[20] E. Kaznowska, J. Depciuch, K. Łach, M. Kołodziej, A. Koziorowska, J. Vongsvivut, I. Zawlik, M. Cholewa, J. Cebulski, The classification of lung cancers and their degree of malignancy by FTIR, PCA-LDA analysis, and a physics-based computational model, Talanta 186 (2018) 337–345.
[21] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[22] Q. Wang, L. Chang, M. Zhou, Q. Li, H. Liu, F. Guo, A spectral and morphologic method for white blood cell classification, Opt Laser. Technol. 84 (2016) 144–148.
[23] X. Zhan, Z. You, C. Yu, L. Li, J. Pan, Ensemble Learning Prediction of Drug-Target Interactions Using GIST Descriptor Extracted from PSSM-Based Evolutionary Information, Biomed Res Int. 11 (2022) 995.
[24] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theor. 13 (1967) 21–27.
[25] M. Sharma, S. Kumar Singh, P. Agrawal, V. Madaan, Classification of clinical dataset of cervical cancer using KNN, India J. Sci. Technol. 9 (2016).
[26] L. Breiman, J.H. Friedman, R.A. Olshen, C.J.J.B. Stone, Classification Regression Tress 40 (1984) 358.
[27] N. Dong, M.-d. Zhai, L. Zhao, C.H. Wu, Cervical cell classification based on the CART feature selection algorithm, J. Ambient Intell. Hum. Comput. 12 (2020) 1837–1849.
[28] J. Laosai, K. Chamnongthai, 2014 International Electrical Engineering Congress, iEECON), 2014.
[29] Y. Xie, H. Yang, X. Yuan, R. Zhang, Q. Zhu, Z. Chu, C. Yang, P. Qin, C. Yan, Stroke prediction from electrocardiograms by deep neural network, Multimed Tools Appl. 80 (2021) 17291–17297.
[30] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, Q. Sun, Deep learning for image-based cancer detection and diagnosis − A survey, Pattern Recogn. 83 (2018) 134–149.
[31] M. Rubin, O. Stein, N.A. Turko, Y. Nygate, N.T. Shaked, TOP-GAN: stain-free cancer cell classification using deep learning with a small training set, Med. Image Anal. 57 (2019) 176–185.
[32] L. Alzubaidi, M.A. Fadhel, O. Al-Shamma, J. Zhang, Y. Duan, Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis, Electronics 9 (2020).
[33] H. Lei, T. Han, F. Zhou, Z. Yu, J. Qin, A. Elazab, B. Lei, A deeply supervised residual network for HEp-2 cell classification via cross-modal transfer learning, Pattern Recogn. 79 (2018) 290–302.
[34] A. Shpilman, D. Boikiy, M. Polyakova, D. Kudenko, A. Burakov, E. Nadezhdina, IEEE International Conference on Machine Learning and Applications, ICMLA, 2017 paper presented at the 2017 16th.
[35] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, N. Houlsby, An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale, 2020.
[36] I. Bello, B. Zoph, A. Vaswani, J. Shlens, Q.V. Le, Attention Augmented Convolutional Networks, 2019.
[37] A. Srinivas, T.Y. Lin, N. Parmar, J. Shlens, A. Vaswani, Bottleneck Transformers for Visual Recognition, 2021.
[38] A. Acevedo, A. Merino, S. Alférez, N. Molina, L. Boldú, J. Rodellar, A dataset of microscopic peripheral blood cell images for development of automatic recognition systems, Data Brief 30 (2020), 105474.

[39] E. Piaton, M. Fabre, I. Goubin-Versini, Recommandations techniques et règles de bonne pratique pour la coloration de May-Grünwald-Giemsa : revue de la littérature et apport de l'assurance qualité, Ann. Pathol. 35 (2015) 294–305.
[40] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems, 1, 2012, pp. 1097–1105.
[41] G.A. Kolokolnikov, A.V. J.t. Samorodov, Comparative Study of Data Augmentation Strategies for White Blood Cells Classification, 2019, pp. 168–175.
[42] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, European Conference on Computer Vision, 2018.
[43] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 5987–5995.
[44] G. Huang, Z. Liu, V. Laurens, K.Q. Weinberger, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 4700–4708.
[45] A.G. Howard, et al., MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017.
[46] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 770–778.
[47] H. Jie, S. Li, S. Gang, S. Albanie, Squeeze-and-Excitation Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
[48] A. Vaswani, N. Shazzer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, arXiv, 2017.
[49] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, J. Shlens, Stand-alone Self-Attention in Vision Models, NeurIPS, 2019.
[50] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding Transfer Learning for Medical Imaging, 2019.
[51] Y. Liu, et al., Mixed-UNet: Refined Class Activation Mapping for Weakly-Supervised Semantic Segmentation with Multi-Scale Inference, 2022, 04227 abs/2205.
[52] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, in: IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 618–626.
[53] A. Acevedo, S. Alférez, A. Merino, L. Puigví, J. Rodellar, Recognition of peripheral blood cell images using convolutional neural networks, Comput. Methods Progr. Biomed. 180 (2019), 105020.
[54] Ucar., Deep learning approach to cell classification in human peripheral blood, in: 2020 5th International Conference on Computer Science and Engineering (UBMK), 2020, pp. 383–387.
[55] F. Long, J.J. Peng, W. Song, X. Xia, J. Sang, BloodCaps: a capsule network based model for the multiclassification of human peripheral blood cells, Comput. Methods Progr. Biomed. 202 (2021), 105972.
[56] M.-E. Percival, C. Lai, E. Estey, C.S. Hourigan, Bone marrow evaluation for diagnosis and monitoring of acute myeloid leukemia, Blood Rev. 31 (2017) 185–192.
[57] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. (2014) 1–42.
[58] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, 2016.