



## A classification method to classify bone marrow cells with class imbalance problem



Liang Guo <sup>a,b,1</sup>, Peiduo Huang <sup>a,1</sup>, Dehao Huang <sup>a</sup>, Zilan Li <sup>a</sup>, Chenglong She <sup>a</sup>, Qianhang Guo <sup>a</sup>, Qingmao Zhang <sup>a</sup>, Jiaming Li <sup>a</sup>, Qiongxiong Ma <sup>a,\*</sup>, Jie Li <sup>c,\*</sup>

<sup>a</sup> Guangdong Provincial Key Laboratory of Nanophotonic Functional Materials and Devices, School of Information and Opto-Electronic Science and Engineering, South China Normal University, Guangzhou 510006, China

<sup>b</sup> Guangdong Provincial Key Laboratory of Industrial Ultrashort Pulse Laser Technology, Shenzhen 518055, China

<sup>c</sup> Department of Hematology, Nan Fang Hospital, Southern Medical University, Guangzhou 510515, China

### ARTICLE INFO

#### Keywords:

Deep learning  
Bone marrow cell classification  
Class-balanced method  
Long-tail distribution

### ABSTRACT

Bone marrow cell morphology has long been used to diagnose blood diseases. However, it requires long-term experience from a suitable person. Furthermore, the outcomes of their recognition are subjective and no quantitative standard has been established yet. Consequently, developing a deep learning automatic system for classifying bone marrow cells is extremely important. However, real-life data sets, such as bone marrow cell data, constantly suffer from a long-tail distribution problem, owing to which the final trained classifier is biased toward a large number of categories. Thus, addressing this issue is crucial. The current research presents a class balance classification method (CBCM) for classifying 15 types of bone marrow cell data sets with a class imbalance problem. CBCM outperforms other balance approaches such as random over-sampling, synthetic minority over-sampling technique (SMOTE), random under-sampling, weighted random forest and weighted cross-entropy function, achieving precision, sensitivity, and specificity values of 84.53%, 84.44% and 99.29% respectively. A more extensive comparison between the baseline and CBCM, as well as the Grad-CAM and Guided Grad-CAM of CBCM, reveals that CBCM is a reliable and effective solution to address the long-tail distribution problem of the bone marrow cell data sets.

### 1. Introduction

For many years, bone marrow aspiration and biopsy have been the universal standard techniques for diagnosing hematological diseases. This method can be used for diagnosing various conditions, including anemia, leukopenia, leukocytosis and pancytopenia, as well as malignant disorders of the blood or bone marrow, such as leukemia, lymphoma and multiple myeloma (MM), along with fever owing to an unknown cause [1]. Although numerous new molecular markers and prognostic techniques have been developed, bone marrow aspiration morphology remains an essential tool for disease diagnosis. Bone marrow specimens must be collected by experienced hematologists, who must then stain and analyze them. However, a skilled hematologist requires several years of training [2,3].

Furthermore, subjective assessments of different people influence the recognition outcome because no quantitative standard has been

established for describing the cell morphology yet [4]. Fig. 1 shows images of bone marrow cells. Actually, for some biological experiments, such as performing electroencephalography (EEG) source analysis [33], it shows the large inter-subject variability over age. It is hard to modelling and it's even more challenging to recognize bone marrow cell images. Various pH values of the bone marrow smear-staining agent, ambient light, and camera conditions yield different images of the same bone marrow cell. Therefore, developing an objective and automated classification method is critical. Several deep learning models [5–8] have been developed as deep learning technology and accompanying hardware facilities have improved, and it has been proven that deep learning outperforms traditional machine learning [9]. Many of them were applied to cell identification [32]. For example, Acevedo [9] suggested a classification approach that employs a convolutional neural network (CNN) trained to identify between eight groups of circulating cells in the peripheral blood. Chuanhao Zhang [10] proposed a novel

\* Corresponding authors.

E-mail addresses: [maqx@m.scnu.edu.cn](mailto:maqx@m.scnu.edu.cn) (Q. Ma), [ljean@163.com](mailto:ljean@163.com) (J. Li).

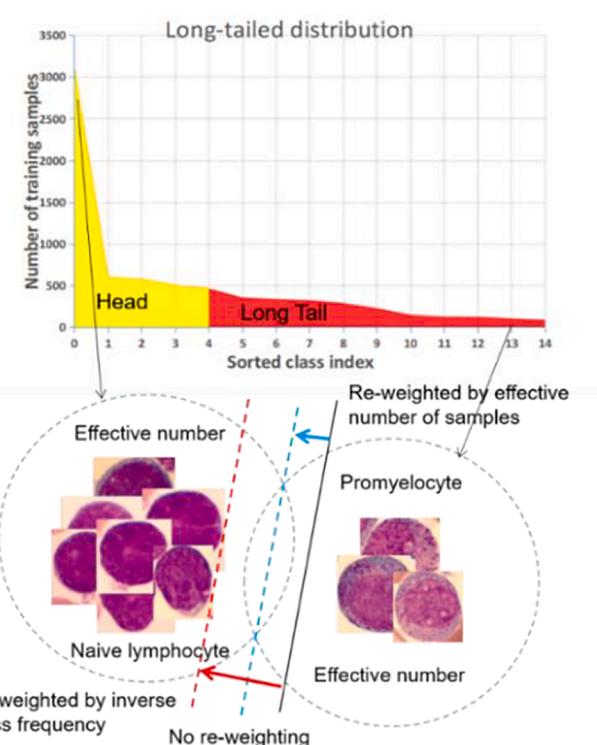
<sup>1</sup> Both authors contributed equally to this work.

hybrid adversarial discriminative network for classifying microscopic leukocyte images.

Although deep learning has achieved good results in blood cell recognition, the class imbalance problem, also known as the long-tailed data distribution problem [11], persist. Some cells, such as proto naive lymphocytes, are easy to collect the sample photos for, while others, such as monocytes, can be obtained only a few samples. Because most learners prefer to recognize the majority class, and in extreme cases, the minority class will be ignored, this dilemma will result in more challenging training. In addition, it frequently misidentifies cells that are few yet critical for disease diagnosis. Acevedo [9] demonstrated that the model's performance is dependent on the balance of the training set of blood cells. Therefore, achieving an effective classification with imbalanced data is crucial [12]. Data-level [13,14] and algorithm-level methods [15–17] have been proposed to handle the class imbalance problem. Data-level methods primarily alter the training data distribution, whereas algorithm-level methods alter the learning or decision process by increasing the importance of the positive class [12].

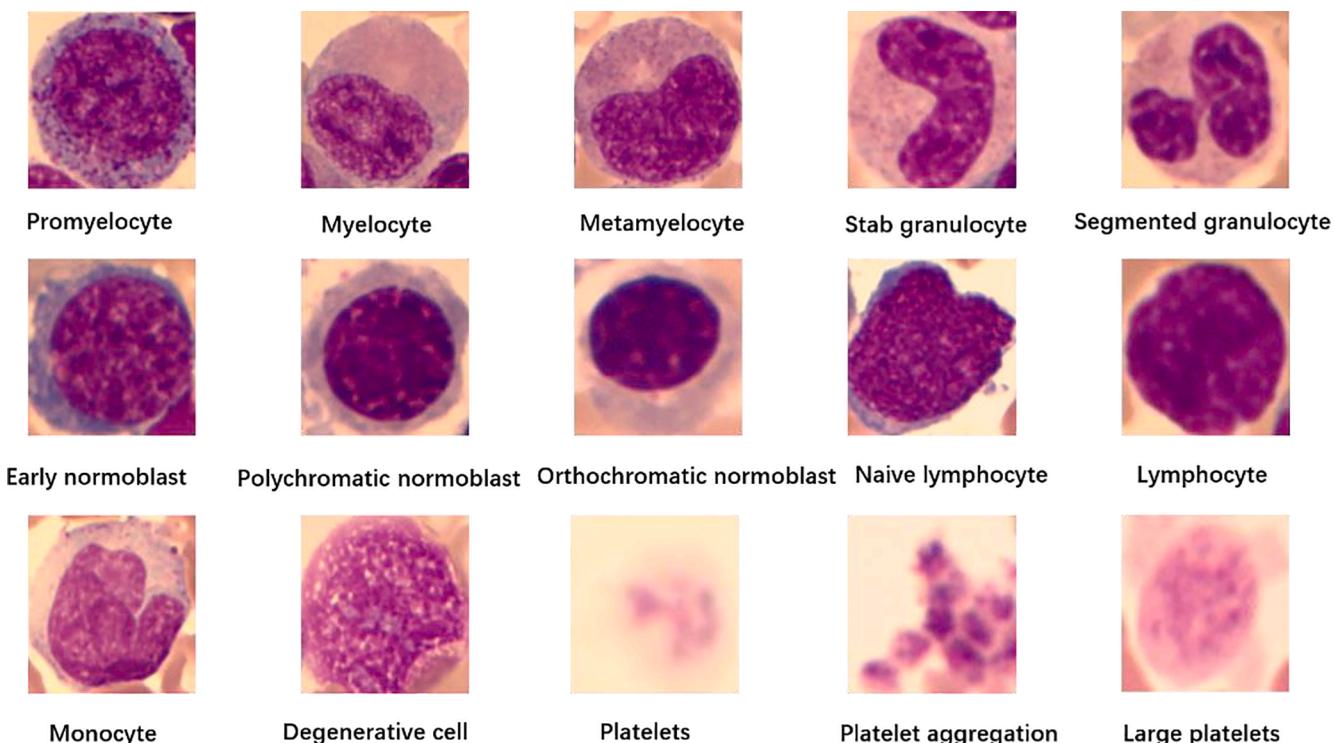
Hence, it is critical to determine effective strategies to overcome the long-tailed data distribution problem in a data set of bone marrow cells. This research presents a class balance classification method (CBCM) for classifying 15 types of bone marrow cell data sets with a class imbalance problem. This method is an algorithm-level method. As observed in Fig. 2, the number of samples in the two groups, i.e., the head and tail of the bone marrow cell data set, differ considerably. We select naive lymphocytes as the head class and promyelocytes as the tail class. The solid black line will reflect the bias toward the dominant class in models without class balance methods trained on these samples. According to inverse class frequency re-weighting [16], this will result in poor performance on the actual data set with a high unbalanced rate, as indicated by the red dashed line. The adequate number of samples [11] is quantified by considering data overlap. A class balance term is designed to reweigh losses by inverting a sufficient number of samples to improve this problem. When trained with CBCM, it can progress the categorization of unbalanced data sets, as indicated by the blue dashed line.

The notable contributions of this study are listed. (1) The class-



**Fig. 2.** Diagram of re-weighting using different techniques.

balanced focal loss (FL) is a major component of CBCM. Compared with the results reported in the literature, the average precision, sensitivity, and specificity of our method are superior to those of other class balance methods. (2) A more extensive comparison of the baseline and CBCM shows that most classes exhibit improved precision and sensitivity. By contrast, almost all categories remain nearly the same or show



**Fig. 1.** Images of bone marrow cells.

slight improvements in terms of specificity. CBCM proved to be beneficial in the classification of such long-tail classes. In other words, in the bone marrow cell data set, CBCM effectively resolves such a class imbalance problem. (3) The Grad-CAM and Guided Grad-CAM of CBCM reveals that CBCM focuses on the cell rather than the environment around the cell; thus, specific fine-grained points can be determined, such as the nucleus and its shape for some cells including stab granulocytes. This feature is similar to the assessment of doctors. This demonstrates that CBCM is reliable.

The rest of the study is organized as follows. Section 2 discusses the materials and methods used in the study, including the data set and the overall structure and components of CBCM. Section 3 discusses the results and presents the experimental charts. Discussion as well as limitations and outlook are described in Section 4. Finally, Section 5 summarizes the important findings of this study.

## 2. Materials and methods

This section is mainly divided into three parts. First, the data set used in this study is introduced. Then, the overall structure of CBCM is described. Finally, the three components of CBCM are presented in detail.

### 2.1. Data set

Southern Hospital of Southern Medical University collected the data set through the optical microscope and camera. Hospital only provided bone marrow images and corresponding labels and did not provide any patient information and original samples. Our research object is bone marrow cell images. Therefore, there is no ethical statement in this paper. The data set contained a total of 7484 images of bone marrow cells. The objective lens is 10 times, the eyepiece is 100 times, and the total magnification is 1000 times (oil lens) to observe the cell morphology. The image is captured by the camera (RGB 1024 × 648 pixels). The operation process is the same as that of the blood smear cell morphology inspection. In the early stage, the doctor extracted the bone marrow of patients through artificial bone marrow puncture and made smears stained. Then the doctor marked the location and category of the cells on the image and finally got a series of single-cell images. All images were resized to 224 × 224 pixels. The data set, a total of 15 categories, respectively is naive lymphocytes, degenerative cells, polychromatic normoblasts, stab granulocytes, metamyelocytes, segmented granulocytes, lymphocytes, myelocytes, orthochromatic normoblasts, early normoblasts, platelet aggregations, monocytes, large platelets, promyelocytes, platelets. Fig. 3 shows the distribution of the data set in each category. From Fig. 3, we can find the number of naive lymphocytes is highest, up to 3097, and the minimum sample is

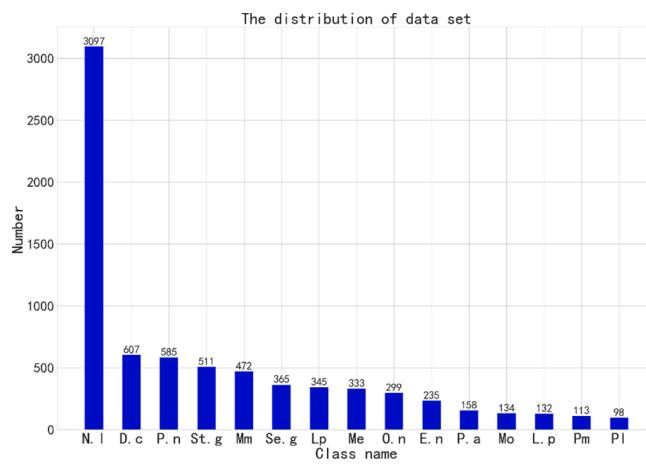


Fig. 3. The distribution of data set.

platelets, which only has 98 images. Thus, the imbalance ratio reached 31:1. The number of training, validation, and test data sets for each category is shown in Table 1.

### 2.2. Overall structure of the proposed classification method

CBCM consists of three main components. The first component uses data pre-processing [18], which can enrich our training samples and reduce overfitting [19,20]. For the second component, pretrained models are used for fine-tuning [21,22], owing to which the model does not need to learn from scratch as most networks. The final component is class-balanced focal loss. The bone marrow cell images are provided as inputs to the system, and the output comprises the correct label indicating the category of each cell image. Fig. 4 shows its overall structure.

### 2.3. Data pre-processing

The pre-processing step plays an essential role in eliminating all kinds of noise in tissue images [18]. CNN needs extensive data set if it wants to get better performance in various evaluation indicators. In the CBCM, a data augmentation technique is applied to increase the data set and reduce the overfitting problems [19,20]. In the data augmentation method, the number of samples increases by applying geometric transformations to the image data sets using simple image processing techniques. The image data set is increased using ColorJitter (randomly changing the brightness, contrast and saturation of an image), RandomHorizontalFlip (horizontally flipping the given Python Image Library (PIL) image randomly with a given probability), RandomVerticalFlip (vertically flipping the given PIL image randomly with a given probability), RandomRotation (rotating the image by random angle) [18]. Examples of some of the transformations applied to augment the number of training images are shown in Fig. 5.

### 2.4. Transfer learning using fine-tuning

As the name suggests, transfer learning is to transfer the learned model parameters to the new model to help the further model training. Considering that most of the data or tasks are relevant, we can transfer the model parameters that have been learned, which can also be understood as the knowledge learned by the model, to the new model in a way to speed up and optimize through transfer learning. Thus, the learning efficiency of the model does not need to learn from scratch like most networks [21,22]. In the experiment, through transfer learning, we used the CNN architectures trained on ImageNet which are called pretrained models to fine-tune the network. CBCM chooses ResNet152 as the baseline model finally and Fig. 6 is the basic network structure of

Table 1  
Number of training, validation, and test data set for each category.

Cell type (abbreviation)	Train data set	Validation data set	Test data set
Naive lymphocytes (N.l)	1841	635	621
Degenerative cells (D.c)	366	122	119
Polychromatic normoblasts (P. n)	335	116	134
Stab granulocytes (St.g)	305	115	91
Metamyelocytes (Mm)	292	89	91
Segmented granulocytes (Se.g)	227	67	71
Lymphocytes (Lp)	194	81	70
Myelocytes (Me)	213	58	62
Orthochromatic normoblasts (O.n)	180	58	61
Early normoblasts (E.n)	145	32	58
Platelet aggregations (P.a)	95	32	31
Monocytes (Mo)	75	30	29
Large platelets (L.p)	93	19	20
Promyelocytes (Pm)	68	24	21
Platelets (Pl)	61	19	18

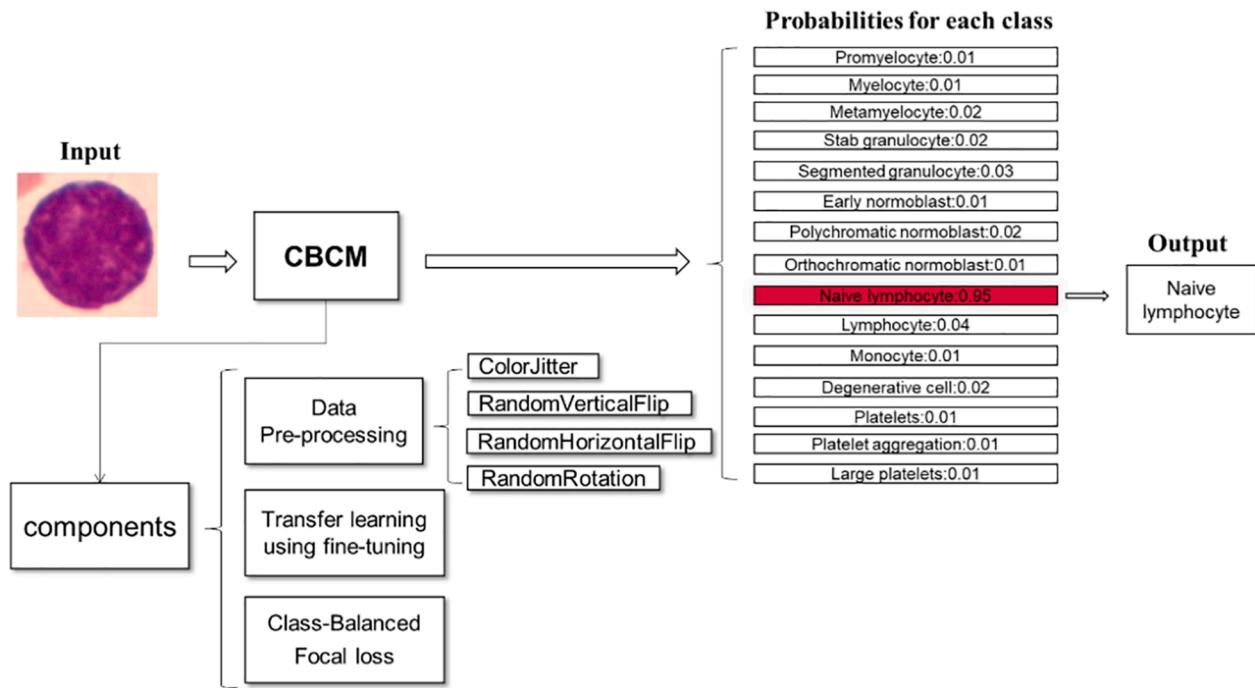


Fig. 4. The overall structure of CBCM.

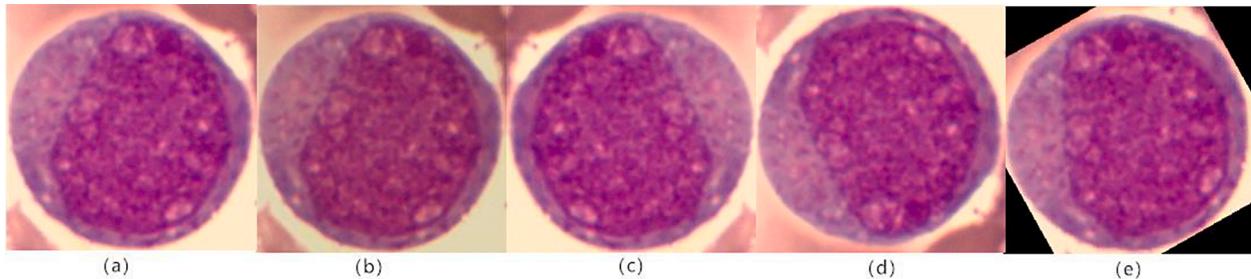


Fig. 5. Promyelocyte: (a) original image, (b) image after applying ColorJitter, (c) image after applying RandomHorizontalFlip, (d) image after applying RandomVerticalFlip, (e) image after applying RandomRotation.

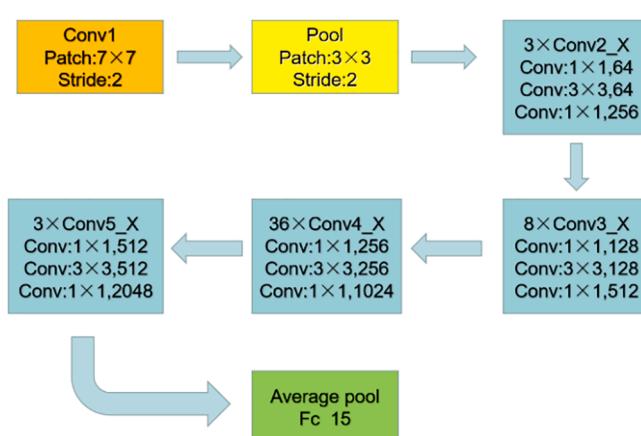


Fig. 6. Basic architecture of ResNet152 using fine-tuning [8].

ResNet152.

### 2.5. Class-balanced focal loss

After an image is classified, the differences between the true label assigned by doctors and the label predicted by the network can be measured by a loss function. For an input sample  $x$  with a label  $y \in \{1, 2, \dots, C\}$ , where  $C$  is the total number of classes. Assume that the output predicted using the fully connected layers of model for all classes is  $z = [z_1, z_2, \dots, z_c]^T$ ; then, we define  $z_i^t$  as:

$$z_i^t = \begin{cases} z_i, & \text{if } i = y. \\ -z_i, & \text{otherwise.} \end{cases} \quad (1)$$

Here,  $\forall i \in \{1, 2, \dots, C\}$ . T.-Y. Lin [23] proposed focal loss (FL). Denote  $p_i^t = \text{sigmoid}(z_i^t) = 1/(1 + \exp(-z_i^t))$ , where  $p_i^t$  presents the probability of being classified to  $i$ . With tunable focusing parameter  $\gamma \geq 0$ , FL can be written as:

$$FL(z, y) = - \sum_{i=1}^C (1 - p_i^t)^\gamma \log(p_i^t) \quad (2)$$

where  $z$  is the output predicted using the fully connected layers of the model and  $y$  is the true class of an input sample. The FL adds a modu-

lating factor  $(1 - p_i^t)^\gamma$  to the sigmoid cross-entropy loss, which reduces the relative loss for well-classified samples and focuses on difficult samples. If  $p_i^t$  given by the classifier is high,  $(1 - p_i^t)^\gamma$  is small, implying that the loss weight of the samples with the correct classification is small. However, the loss weight of the samples with the wrong classification is large. In practice, it uses an  $\alpha$ -balanced variant of the FL:

$$FL(z, y) = -\alpha_t \sum_{i=1}^C (1 - p_i^t)^\gamma \log(p_i^t) \quad (3)$$

Yin Cui [11] proposed a concept of the effective number of samples. All augmented examples such as random cropping, re-scaling and horizontal flipping for CNNs are considered the same as original examples. Therefore, denote the effective number (expected volume) of samples as  $E_n$ , where  $n \in \mathbb{Z} > 0$  is the number of samples. The mathematical formulation is shown below:

$$E_n = (1 - \beta^{n_y}) / (1 - \beta) \quad (4)$$

After determining the effective number of samples, a weighting factor  $\alpha_t$  can be obtained, where  $\alpha_t = 1/E_n = (1 - \beta)/(1 - \beta^{n_y})$  and  $n_y$  presents the effective number of the true class  $y$ , which can be applied to class-balanced FL. Further, the class-balanced (CB) FL is expressed as:

$$CB_{focal}(z, y) = -\frac{1 - \beta}{1 - \beta^{n_y}} \sum_{i=1}^C (1 - p_i^t)^\gamma \log(p_i^t) \quad (5)$$

Therefore, the parameter  $\gamma$  can solve the imbalance problem between difficult and easy samples, thereby ensuring that the classifier focuses on difficult samples and ignores the easy ones. The parameter  $\beta$  solves the class imbalance problem through the effective number [11]. For parameter selection, the parameter  $\gamma$  mainly selects values from {0.5, 1.0, 2.0} and parameter  $\beta$  mainly selects values from {0.9, 0.99, 0.999, 0.9999} [11,23].

### 3. Results

This section shows the results of three experiments and some analysis of results. In the first experiment, a comparison is performed between traditional machine learning and CNN. Moreover, this experiment is performed to identify the need to use data pre-processing. In the second experiment, the most suitable CNN architecture is determined. In the third experiment, the most suitable approach for solving the long-tailed data distribution problem based on the selected baseline is determined. The evaluation indicators used in this paper are mainly precision, sensitivity, specificity [9]. All experiments are conducted on an Ubuntu system and the networks were trained using Pytorch software libraries [18] and an Nvidia Titan XP GPU.

#### 3.1. Comparison between traditional machine learning and CNN

In this experiment, traditional machine learning approaches and CNN are compared. Moreover, this experiment is performed to identify the need to use data pre-processing. Traditional machine learning approaches include support vector machine (SVM) [36], and random forest [37]. The CNN approach is ResNet152. The loss function used in this

experiment is cross-entropy. Table 2 presents a comparison between traditional machine learning and deep learning approaches. As observed from this table, ResNet152 outperforms traditional machine learning approaches. After using data pre-processing, the performance of the deep learning approach improves. A considerable improvement is observed in precision, sensitivity, and specificity, which increase by 2.27%, 3.5% and 0.16% respectively. According to the obtained data, CNN is better suited for classifying bone marrow cell images and data pre-processing is advantageous to CNN.

#### 3.2. Results of selecting CNN architecture

In this experiment, fine-tuning is performed to determine the most suitable CNN architecture and compare this architecture with those of other methods that also achieved blood cell recognition using deep learning. Each method has its own CNN architectures, and the loss function used in this experiment is cross-entropy [18]. All CNN architectures use data pre-processing. Table 3 shows the results of different CNN architectures. As observed from this table, ResNet152 achieves the performance in terms of all three evaluation indicators. A considerable improvement is observed in precision, sensitivity, and specificity, [9] which increase by 4.27%, 7.31%, and 0.4% respectively, compared with the approach proposed by Sarmad Shafique [24]. Among them, the sensitivity of ResNet152 is more prominent than the other three methods. Finally, based on these findings, ResNet152 is selected as the baseline model.

#### 3.3. Results of different methods to solve the long-tailed data distribution problem

In this experiment, different class balance methods are compared and a method that is most suitable for the bone marrow cell data set is determined. For comparison, random over-sampling [26], synthetic minority oversampling technique (SMOTE) [34], random under-sampling [27], weighted random forest [31], weighted cross-entropy function [28], and CBCM are considered. All methods use data pre-processing. For SMOTE and weighted random forest, the features were extracted automatically by baseline model. Table 4 shows the results of different class balance methods. CBCM performs best in terms of precision, sensitivity, and specificity. There are noticeable improvements in precision and sensitivity, which increase by 1.62% and 1.88% respectively, while the improvement in specificity is less obvious when using CBCM. This shows that CBCM is a viable method for addressing the long-tail distribution problem in a blood cell data set. Furthermore, in this experiment,  $\gamma = 2.0$  and  $\beta = 0.9$  (equation (5)).

Based on the results of this experiments, it can be concluded that CBCM outperforms other class balance methods with respect to the bone marrow cell data set. Thus, it is an effective method for addressing the issue of long-tail distribution of bone marrow cells.

**Table 3**  
Different CNN architectures tested results.

Methods	CNN architectures	Precision	Sensitivity	Specificity
Sarmad Shafique et al. (2018) [24]	Alexnet	0.7864 ± 0.0103	0.7525 ± 0.0130	0.9877 ± 0.0006
Andrea Acevedo et al. (2019) [9]	Vgg16	0.8254 ± 0.0060	0.8111 ± 0.0147	0.9909 ± 0.0004
Feiwei Qin et al. (2018) [25]	Googlenet	0.8251 ± 0.0099	0.8115 ± 0.0091	0.9914 ± 0.0003
ResNet152	ResNet152	0.8291 ± 0.0061	0.8256 ± 0.0062	0.9917 ± 0.0005

**Table 2**  
Comparison between traditional machine learning and deep learning.

Approaches	Precision	Sensitivity	Specificity
Support Vector Machine [34]	0.5343 ± 0.0000	0.4215 ± 0.0000	0.9504 ± 0.0000
Random Forest [35]	0.5337 ± 0.0146	0.4200 ± 0.0107	0.9684 ± 0.0005
Resnet152	0.8064 ± 0.0113	0.7906 ± 0.0064	0.9901 ± 0.0003
Resnet152 + data pre-processing	0.8291 ± 0.0061	0.8256 ± 0.0062	0.9917 ± 0.0005

**Table 4**

Different class balance methods tested results.

Methods	Precision	Sensitivity	Specificity
Baseline	0.8291 ± 0.0061	0.8256 ± 0.0062	0.9917 ± 0.0005
Random over-sample [26]	0.8396 ± 0.0078	0.8263 ± 0.0095	0.9922 ± 0.0004
SMOTE [13]	0.8279 ± 0.0047	0.8302 ± 0.0029	0.9922 ± 0.0002
Random under-sample [27]	0.7563 ± 0.0108	0.7994 ± 0.0203	0.9877 ± 0.0005
Weighted random forest [31]	0.8314 ± 0.0064	0.8174 ± 0.0050	0.9916 ± 0.0003
Weighted cross-entropy function [28]	0.8327 ± 0.0044	0.8375 ± 0.0054	0.9921 ± 0.0002
CBCM	<b>0.8453 ± 0.0022</b>	<b>0.8444 ± 0.0072</b>	<b>0.9929 ± 0.0002</b>

### 3.4. Detailed analysis of the confusion matrix

Fig. 7 shows two confusion matrices of the baseline and CBCM test assessments. The differences between the two confusion matrices can be found in Fig. 7 (a) and (b). First, the number of correctly identified cells, naive lymphocytes, which has the highest number of images in the training data set, decreased from 610 to 604 after using CBCM. However, other cells increased, such as promyelocytes, myelocytes, metamyelocytes, stab granulocytes, segmented granulocytes, early normoblasts, polychromatic normoblasts, lymphocytes, monocytes, degenerative cells, and platelet aggregation, thereby increasing the overall accuracy from 89.5% to 90.9%. Furthermore, precision and sensitivity improved considerably.

To analyze indicators of each class, bars presenting the average precision, sensitivity, and specificity for each class were plotted (Fig. 8). The bar chart observation and analysis reveal a significant improvement in precision and sensitivity for most cells, including promyelocytes, metamyelocytes, stab granulocytes, polychromatic normoblasts, degenerative cells, and large platelets. Furthermore, the specificity of almost all types of cells remained approximately the same or improved slightly. These findings indicate that CBCM is effective in resolving such class imbalance problems in the bone marrow cell data set.

Additionally, based on the two confusion matrices, after using CBCM, many types of cells are correctly or incorrectly recognized as cells in adjacent stages. For example, for polychromatic normoblasts, the most incorrectly recognized cells are early and orthochromatic normoblasts, which are the adjacent stages of the cell. The reason for this phenomenon may be related to the data set marked by the doctors. Cell morphology is very similar in adjacent stages, and the cell cycle is rapid. Consequently, when the assessment of cells with identical cell characteristics becomes complex, doctors classify this cell as being in the latter stage.

### 3.5. Grad-CAM and Guided Grad-CAM of the proposed classification method

Grad-CAM uses gradient information flowing into the last convolutional layer of the CNN to understand the importance of each neuron for a decision of interest. Although our technique is very generic and can be used to visualize any activation in a deep network, we focus on explaining the model's decisions in this study. Although Grad-CAM visualizations are class discriminative and localize relevant image regions well, they cannot show fine-grained features as pixel-space gradient visualization methods, such as guided backpropagation [30]. To combine the best aspects of both, we fuse guided backpropagation and Grad-CAM visualizations using point-wise multiplication called Guided Grad-CAM [29].

Fig. 9 shows the original, Grad-CAM, and Guided Grad-CAM images of stab granulocyte cells. In the Grad-CAM image, the classifier localizes

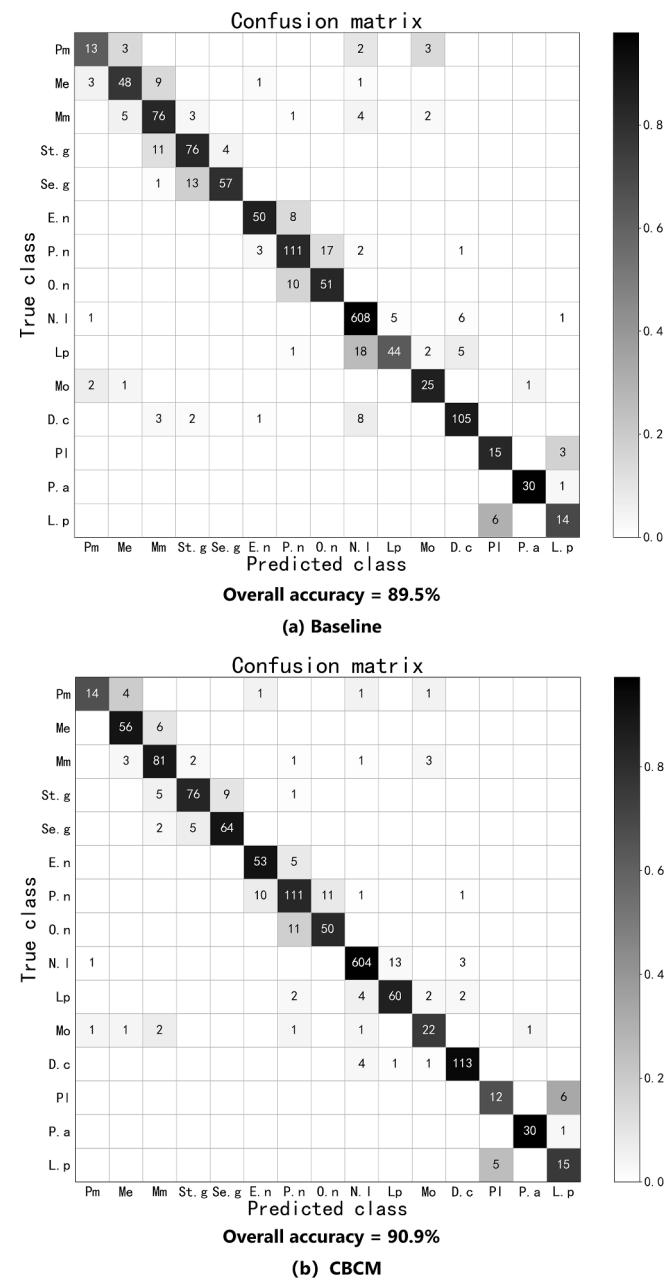
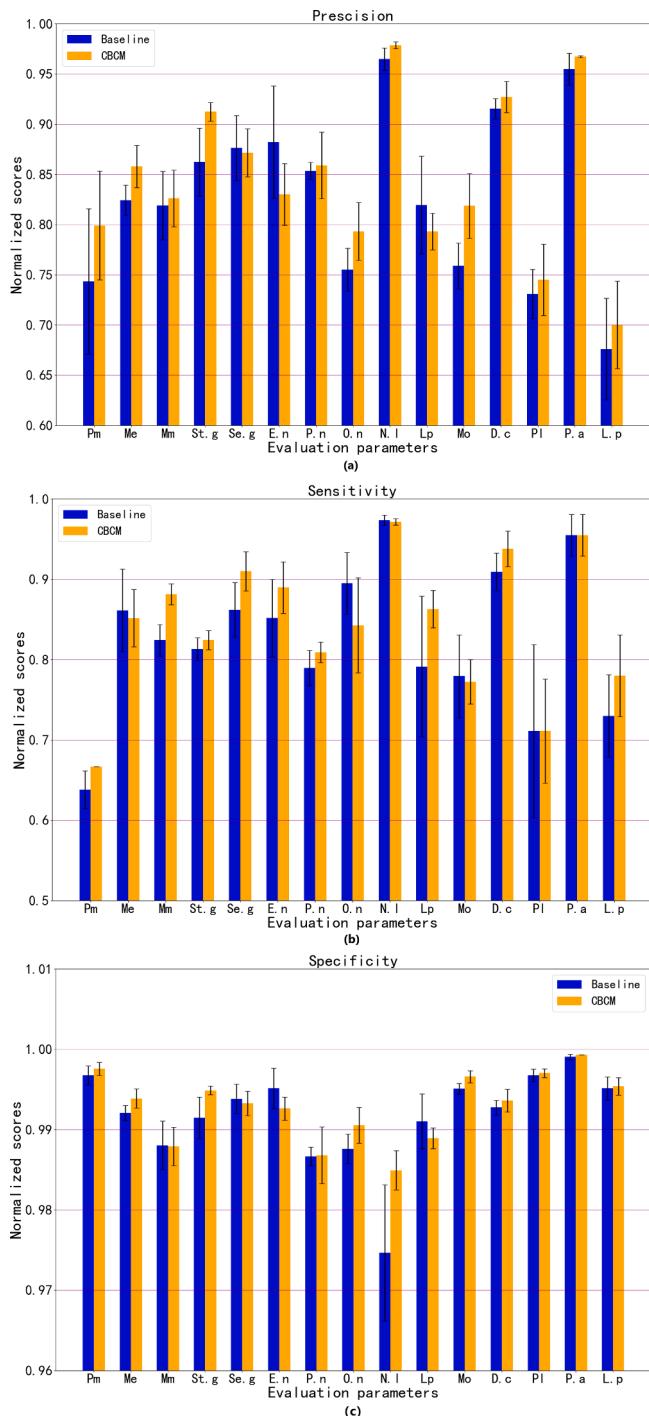


Fig. 7. Confusion matrices of the test assessment of the baseline and CBCM. Rows represent the true values and columns represent the predicted values. The principal diagonal (gray) is the number of correctly recognized classes.

the area whose color is red. Thus, it focuses on the cell instead of the environment around the cell. In the Guided Grad-CAM image, more fine-grained points are focused on, such as the nucleus and its shape. A noticeable feature of stab granulocyte cells is their rod-shaped nuclei, which are consistent with the fine-grained points that the classifier focuses on. This demonstrates the reliability of the proposed model.

## 4. Discussion

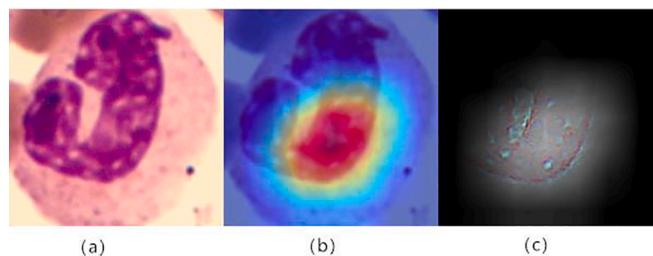
At present, the recognition of blood cells basically stays in a few categories and few researchers study more than 10 types. Acevedo [9] suggested a classification approach that employs a convolutional neural network (CNN) trained to identify between eight groups of circulating cells in the peripheral blood. Muhammed Yildirim [36] used CNN to classify four types of white blood cells. Roopa B. Hegde [37]



**Fig. 8.** (a) Average precision of each class based on the baseline and CBCM, (b) average sensitivity of each class based on the baseline and CBCM, and (c) average specificity of each class based on the baseline and CBCM.

demonstrated the classification of six types of white blood cells using traditional image processing approach and deep learning approach. However, when studying over 10 types cell classification, the class imbalance problem seemed obvious.

In our research, we focused on classification of bone marrow cells into fifteen types, which is more fine-grained. However, due to the class imbalance problem, the classifier would tend to the category with large amount. Besides, in some extreme cases, the minority class will be ignored, this dilemma will result in bad model training. In addition, it frequently misidentifies cells that are few yet critical for disease



**Fig. 9.** Stab granulocyte (a) original image, (b) Grad-CAM image, (c) Guided Grad-CAM image.

diagnosis. Therefore, we proposed a classification method CBCM to classify the 15 types of bone marrow cell data with imbalance problem.

CBCM is based on Resnet152. It has a better performance than traditional machine learning and some deep learning models, such as Alexnet, Vgg16, Googlenet [9,24,25,34,35]. Because the core of ResNet is to solve the side effect (degradation problem) caused by increasing the depth of the network, which can improve the network performance by simply increasing the depth of the network.

Few researchers have studied the imbalance problem of bone marrow cells. Even if there is, they just stayed in the discovery and did not specifically put forward a better method to solve it. For example, Acevedo [9] found that the balanced data was better for identification. The finding seemed consistent with our study. But it did not propose how to solve the imbalanced data. Thus, this paper proposes CBCM to solve the class imbalance problem in the automatic recognition of 15 types of bone marrow cells, which can provide a reference for researchers who want to study related problems. The main component of CBCM, class-balance FL can consider the class imbalance problem using the effective number and the data imbalance problem between difficult and easy samples. Compared to those data-level methods and other reweighted algorithms [13,26–28,31], it can solve the class imbalance problem more effectively. In this study, CBCM was demonstrated that it is an effective and reliable method to bone marrow cells recognition according to detailed analysis of the Grad-CAM and Guided Grad-CAM of CBCM and confusion matrix.

The purpose of bone marrow cell recognition is to diagnose diseases. When we can recognize more cell types, it is possible to diagnose more detailed diseases (disease subtypes) for better treatment. CBCM can reduce the false identification for minority classes, solve the class imbalance problem, make the automatic identification technology of fifteen types of bone marrow cells more accurate, which can achieve effective diagnose diseases and make it possible to be applied to clinical practice in the future. The remaining subsections discuss some limitations and outlook of the experiments and study.

#### 4.1. Limitations and outlook

There is insufficient automation for parameter  $\gamma$  and  $\beta$  to be adjusted, thereby requiring manual adjustments. Furthermore, a more detailed comparison of the baseline and CBCM results in a few cells being misidentified as cells in adjacent stages after using CBCM.

In future studies, the problem of incorrectly recognizing cells in adjacent stages is supposed to be solved. Besides, work has been carrying out to develop a more effective algorithm for the class imbalance problem, such as focusing on small sample identification. When the class imbalance problem is solved, bone marrow cell recognition can be improved to more categories.

#### 5. Conclusions

This study proposes the CBCM to classify 15 types of bone marrow cell data sets with class imbalance problems. CBCM can solve the class imbalance problem more effectively than other class balance methods.

Compared to previous studies, the above experiments found that the precision, sensitivity, and specificity are better than other class balance methods. Additionally, based on a detailed analysis of the comparison between the baseline and CBCM, a noticeable improvement is observed in terms of precision and sensitivity for most classes, and almost all categories remain roughly the same or show slight improvements in terms of specificity. It is demonstrated that CBCM is useful for classifying long-tail distribution problems. Thus, CBCM is effective in solving such class imbalance problems in the bone marrow cell data set. Finally, using the Grad-CAM and Guided Grad-CAM of CBCM, it is revealed that CBCM focuses on the cell instead of the environment around the cell. For some cells such as stab granulocytes, CBCM focuses on some fine-grained points, such as the nucleus and its shape. This feature is roughly the same as the features assessed by doctors. These findings show that CBCM is reliable.

During this study, it is discovered that the majority of the cells that were incorrectly identified after using the class balance method are concentrated in the adjacent stage of the correct cell, providing a way for subsequent developers to continue to improve the classification system.

#### CRediT authorship contribution statement

**Liang Guo:** Conceptualization, Methodology, Software. **Peiduo Huang:** Methodology, Software, Writing – original draft. **Dehao Huang:** Formal analysis, Data curation. **Zilan Li:** Investigation. **Chenglong She:** Validation. **Qianhang Guo:** Visualization. **Qingmao Zhang:** Supervision. **Jiaming Li:** Writing – review & editing. **Qiongxiong Ma:** Project administration, Writing – review & editing. **Jie Li:** Writing – review & editing.

#### 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.

#### Acknowledgment

This work was supported by Department of Science and Technology of Guangdong Province (2018B030323017); Key-Area Research and Development Program of Guangdong Province (No. 2020B090922006); National Key Research and Development Program of China (2017YFB1104500); National Natural Science Foundation of China (62005081); Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation ("Climbing Program" Special Funds) (No. pdjh2021b0134); Science and Technology Program of Guangzhou (No. 202002030165, 2019050001); Featured Innovation Project of Guangdong Education Department (No. 2019KTSCX034); Young Innovative Talents Project in Universities of Guangdong Province (No. 2018KQNXCX057); Young Scholar Foundation of South China Normal University (No. 19KJ13).

#### References

- [1] D. Focosi, Bone Marrow Aspiration and Biopsy, *N. Engl. J. Med.* 362 (2) (2010) 182–183.
- [2] Wu Yi-Ying, Huang Tzu-Chuan, Ye Ren-Hua, Hematologist - Level Deep Learning Algorithm (BMSNet) for Assessing the Morphologies of Single Nuclear Balls in Bone Marrow Smears: Algorithm Development, *Jmir Med. Inform.* 8 (4) (2020), <https://doi.org/10.2196/15963>.
- [3] S.H. Lee, W.N. Erber, A. Porwit, M. Tomonaga, L.C. Peterson, ICSH guidelines for the standardization of bone marrow specimens and reports, *International Journal of Laboratory Hematology* 30 (5) (2008) 349–364, <https://doi.org/10.1111/j.1751-553X.2008.01100.x>.
- [4] Shi Tingting, Research on Key Techniques of Automatic Examination of Bone Marrow Cell Morphology for Acute Leukemia, Master dissertation, South China University of Technology, 2018.
- [5] Alex Krizhevsky, One weird trick for parallelizing convolutional neural networks, *arXiv: Neural and Evolutionary Computing* (2014).
- [6] Simonyan Karen, Zisserman Andrew, Very Deep Convolutional Networks for Large-Scale Image Recognition, in: *computer vision and pattern recognition* (2014).
- [7] Szegedy C, Wei Liu, Yangqing Jia, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, Going Deeper with Convolutions, in: *2015 IEEE Conference on Computer Vision and Pattern Recognition* (2015), 10.1109/cvpr.2015.7298594.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778, <https://doi.org/10.1109/cvpr.2016.90>.
- [9] A. Acevedo, S. Alferez, A. Merino, L. Puigvi, J. Rodellar, Recognition of peripheral blood cell images using convolutional neural networks, *Comput. Methods Programs Biomed.* 180 (2019), <https://doi.org/10.1016/j.cmpb.2019.105020>.
- [10] Zhang Chuanhao, Wu Shangshang, Lu Zhiming, Hybrid adversarial-discriminative network for leukocyte classification in leukemia, *Medical Physics*, (2020), 10.1002/mp.14144.
- [11] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, Belongie. S, Class-balanced loss based on effective number of samples, in: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019) pp. 9260–9269, 10.1109/cvpr.2019.00949.
- [12] J.M. Johnson, T. Khoshgoftaar, M, Survey on deep learning with class imbalance, *Journal of Big Data* 6 (1) (2019) 27, <https://doi.org/10.1186/s40537-019-0192-5>.
- [13] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Smote:synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.
- [14] Stefanowski, J.: Dealing with data difficulty factors while learning from imbalanced data. In: *Challenges in Computational Statistics and Data Mining*, pp. 333–363 (2016).
- [15] B. Krawczyk, Learning from imbalanced data: open challenges and future directions, *Progress in Artificial Intelligence* 5 (4) (2016) 221–232, <https://doi.org/10.1007/s13748-016-0094-0>.
- [16] Z.-H. Zhou, X.-Y. Liu, On multi-class cost-sensitive learning, *Comput. Intell.* 26 (3) (2010) 232–257, <https://doi.org/10.1111/j.1467-8640.2010.00358.x>.
- [17] Bartosz Krawczyk, Michał Woźniak, Francisco Herrera, On the usefulness of one-class classifier ensembles for decomposition of multi-class problems, *Pattern Recognit.* 48 (12) (2015) 3969–3982.
- [18] <https://pytorch.org>.
- [19] A. Krizhevsky, I. Sutskever, G. Hinton, E, ImageNet Classification with Deep Convolutional Neural Networks, *Commun. ACM* 60 (6) (2017) 84–90, <https://doi.org/10.1145/3065386>.
- [20] D. Ciresan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3642–3649, <https://doi.org/10.1109/cvpr.2012.6248110>.
- [21] L. Yang, S. Hanneke, J. Carbonell, A theory of transfer learning with applications to active learning, *Machine Learning* 90 (2) (2013) 161–189, <https://doi.org/10.1007/s10994-012-5310-y>.
- [22] S. Khan, N. Islam, Z. Jan, I. Din, U, Rodrigues Jjrc, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, *Pattern Recogn. Lett.* 125 (2019) 1–6, <https://doi.org/10.1016/j.patrec.2019.03.022>.
- [23] Tsung-Yi Lin, P. Goyal, R. Girshick, Kaiming He, P. Dollar, Focal loss for dense object detection, in: *2017 IEEE International Conference on Computer Vision*, 2017, pp. 2999–3007, <https://doi.org/10.1109/iccv.2017.324>.
- [24] Shafique Sarmad, Tehsin Samabia, Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks, *Technol. Cancer Res. Treat.* 17 (2018), <https://doi.org/10.1177/1533033818802789>.
- [25] Qin Feiwei, Gao Nannan, Wu Peng Yong, Shen Shuying Zizhao, Grudtsin Artur, Fine-grained leukocyte classification with deep residual learning for microscopic images, *Comput. Methods Programs Biomed.* 162 (2018) 243–252, <https://doi.org/10.1016/j.cmpb.2018.05.024>.
- [26] C.X. Ling, Chenghui Li, Data mining for direct marketing: problems and solutions, in: *Fourth International Conference on Knowledge Discovery and Data Mining*, 1998, pp. 73–79.
- [27] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, Gong Bing, BingGong, Learning from class imbalanced data: Review of methods and applications, *Expert Syst. Appl.* 73 (2017) 220–239, <https://doi.org/10.1016/j.eswa.2016.12.035>.
- [28] Y.S. Aurelio, G.M. de Almeida, C.L. de Castro, A.P. Braga, Learning from Imbalanced Data Sets with Weighted Cross-Entropy Function, *Neural Processing Letters* 50 (2) (2019) 1937–1949, <https://doi.org/10.1007/s11063-018-09977-1>.
- [29] Selvaraju R, R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: *2017 IEEE International Conference on Computer Vision* (2017), pp. 618–626, 10.1109/iccv.2017.74.
- [30] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller. Striving for Simplicity: The All Convolutional Net. *ICLR-2015*, arXiv: 1412.6806, 2014. 2, 3.
- [31] <https://scikit-learn.org>.
- [32] Mishra.S, Majhi.B and Sa.PK. Texture feature based classification on microscopic blood smear for acute lymphoblastic leukemia detection, *Biomedical signal processing and control*, 47(303-311), 2019.
- [33] Marios Antonakakis, Sophie Schrader, Ümit Aydin, Asad Khan, Joachim Gross, Michalis Zervakis, Stefan Rampp, Carsten H. Wolters, Inter-Subject Variability of Skull Conductivity and Thickness in Calibrated Realistic Head Models, *Neuroimage* 223 (2020) 117353, <https://doi.org/10.1016/j.neuroimage.2020.117353>.
- [34] Chang,CC, LIBSVM: A Library for Support Vector Machines.Acm transactions on intelligent systems and technology,2(3),2011, 10.1145/1961189.1961199.

- [35] L. Breiman, Random forests, *Machine learning* 45 (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [36] Muhammed Yildirim, Ahmet Çinar, Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease, *Revue d'Intelligence Artificielle* 33 (5) (2019) 335–340.
- [37] Roopa B. Hegde, Keerthana Prasad, Harishchandra Hebbar, Brij Mohan Kumar Singh, *Engineering* 39 (2) (2019) 382–392.