The Journal of Supercomputing (2024) 80:249–270
https://doi.org/10.1007/s11227-023-05490-y

White blood cell classification based on a novel ensemble convolutional neural network framework

Na Dong[1] · Qingyue Feng[1] · Jianfang Chang[1] · Xiaoming Mai[1]

Accepted: 6 June 2023 / Published online: 23 June 2023

Abstract

White blood cell detection plays an integral role in diagnosing pathologies such as leukemia and gestational diabetes. Despite this, conventional image-based white blood cell classification methodologies encounter obstacles including inaccurate cell segmentation and labor-intensive artificial feature extraction. Contrarily, Convolutional Neural Networks (CNNs) have the capacity to learn features autonomously from raw images, thereby offering a novel and effective solution for blood cell detection. Notwithstanding, the features ascertained by a solitary CNN tend to be unidirectional. Conversely, ensemble learning combines results from numerous networks, thus ensuring an adequate acquisition of feature information and subsequently enhancing the model's overall efficacy. Consequently, this study introduces a method for white blood cell classification underpinned by ensemble CNNs. Initially, three high-performing CNNs possessing disparate structures, namely VGG16, ResNet50, and Inception V3, are enlisted as base learners to augment the diversity of base learners. Subsequently, the Gompertz function is employed to strategize the ensemble learning combination strategy, taking into consideration the prediction confidence and fuzzy level of each base learner. Ultimately, the ensemble CNN model is developed, incorporating learning outcomes from several singular models and utilizing diversified information to achieve white blood cell classification. Empirical results indicate that the ensemble learning technique advanced in this study enables accurate and reliable white blood cell classification, demonstrating potential clinical value.

Keywords White blood cell classification · Convolutional neural network · Ensemble learning · Combination strategy

Extended author information available on the last page of the article

1 Introduction

White blood cells serve as distinctive immune cells, which are integral to preserving the human body's immune functionality. This group encompasses five cell varieties: neutrophils, eosinophils, basophils, monocytes, and lymphocytes. Under pathological circumstances, white blood cell count and morphology may undergo alterations. Detecting quantitative or qualitative shifts in these different types of cells aids physicians in disease diagnosis, making the investigation of white blood cell classification methodologies clinically significant.

Currently, clinical white blood cell categorization chiefly utilizes manual microscopic examination and blood cell analyzers. However, the manual method is labor-intensive, inefficient, and subject to observer expertise, whereas blood cell analyzers are costly and lack the capacity to observe cellular morphological changes. With artificial intelligence's rapid advancement, the trend is toward employing computer image processing technology for automated white blood cell classification. An increasing number of researchers are exploring automated detection techniques for white blood cells, primarily deploying machine learning and deep learning methods. Conventional white blood cell classification methodologies involve cell segmentation, feature extraction, and classifier design.

Li et al. [1] introduced a dual-threshold segmentation method based on RGB and HSV color spaces. In the process of white blood cell segmentation, the background and red blood cells were separated from the original images by two distinct thresholds. Zheng et al. [2] utilized K-means clustering to differentiate background regions and extracted white blood cell regions based on contact cell division via concavity analysis. Pergad et al. [3] extracted statistical features from segmented nuclei and applied the Harr wavelet transform to derive wavelet features from preprocessed images. Tavakoli et al. [4] extracted three shapes and four novel color features from the nucleus and cytoplasm, utilizing Support Vector Machine (SVM) for white blood cell classification. Khan et al. [5] discarded redundant features using a feature selection technique premised on maximum correlation probability and deployed Extreme Learning Machine (ELM) to classify the chosen features. When employing these traditional machine learning methods for white blood cell classification, accurate cell segmentation and effective feature extraction become pivotal. Yet, accurate white blood cell segmentation remains a significant challenge due to intrinsic cellular texture and background interference, coupled with the absence of a standardized protocol for artificial features, which is one-sided and subjective.

Given the swift advancements of deep learning models, the Convolutional Neural Network (CNN) has found extensive applications in diverse image processing domains, such as image classification, semantic segmentation, and object detection [6–8]. In 2012, Krizhevsky et al. created the record of the field of image classification with a novel network called AlexNet [9]. Subsequently, a plethora of exceptional neural network algorithms have emerged, demonstrating impressive performance in image classification tasks. For instance, Jiang et al. [10] proposed an innovative CNN model capable of thoroughly extracting white blood

cell image features by incorporating the batch normalization algorithm, residual convolution architecture, and enhanced activation function, hence facilitating rapid white blood cell identification. Banik et al. [11], aiming to classify localized white blood cell images, devised a novel CNN model by merging the features of the initial and final convolutional layers, allowing the model to train with fewer parameters and computational complexity. Bagido et al. [12] utilized transfer learning with the pre-trained Inception ResNetV2 model, displaying proficient classification performance when distinguishing different white blood cells. Liang et al. [13] proposed a deep neural network architecture for white blood cell detection, deepening image content comprehension by amalgamating the features of the convolutional neural network and Recurrent Neural Network (RNN). In their efforts to automatically identify and classify white blood cell images, Mohamed et al. [14] employed the pre-trained model as the automatic feature extractor, using the traditional machine learning classifier for classification. This model synthesizes the strengths of deep models in automatic feature extraction with superior classification accuracy.

Contrary to traditional machine learning methods, the convolutional neural network directly procures feature information from the original images, circumventing the need for cell segmentation and feature extraction, thus considerably reducing the researchers' workload. While the convolutional neural network possesses robust self-learning capabilities, the feature information derived from a single model may be inadequate, potentially compromising the accuracy of classification outcomes. Ensemble learning leverages model diversity to yield superior results compared to singular models. To enhance the reliability and generalization of white blood cell classification, this paper provides the following primary contributions:

1. This study introduces an innovative white blood cell classification methodology premised on the ensemble convolutional neural network. Employing three exemplary networks, namely VGG16, ResNet50, and Inception V3 as base learners, the classification of white blood cells is realized through the integration of several individual models' learning outcomes.
2. The Gompertz function is incorporated to formulate a combination strategy for ensemble learning. The corresponding fuzzy level is generated using the prediction confidence of each base learner, concurrently considering the influence of prediction confidence and fuzzy level of each base learner on classification results.

The rest of this paper is structured as follows: Sect. 2 delineates the chosen base learners; Sect. 3 outlines the proposed method's design; Sect. 4 engages in a discourse on the experimental results and analysis; and finally, Sect. 5 provides the conclusion and future prospects.

2 Selection of base learners

CNN is a representative model in the realm of deep learning, primarily encompassing convolutional layers, pooling layers, and fully connected layers. In current white blood cell classification and identification, a singular network model is predominantly employed; however, the information derived from this approach is unilateral. To maximize the utilization of white blood cell image features, this study employs three distinct CNN models (VGG16, ResNet50, and Inception V3) as base learners. This approach procures abundant feature information through model diversity, thereby enhancing the proficiency in recognizing white blood cell images. The basic architecture of each CNN model is delineated in the subsequent subsections.

2.1 VGG16 network

The VGGNet, a convolutional neural network model collaboratively designed by Oxford University's Visual Geometry Group and Google DeepMind [15], encompasses various layers of networks with the depth of weight layers ranging from 11 to 19. In this study, the VGG16 model with 16 weight layers is employed. Its network structure is depicted in Fig. 1.

The VGG16 network structure comprises five convolution blocks, each hosting 2 ~ 3 convolution layers with a convolution kernel size of 3 × 3 and a stride of 1. Each convolutional block is succeeded by a max-pooling layer, with a pooling kernel size of 2 × 2 and a stride of 2. The final three layers are fully connected: the initial two layers comprise 4096 channels, and the third encompasses 1000 neurons, symbolizing 1000 categories of the ImageNet dataset. Finally, the Softmax function outputs the prediction result.

The VGG16 model employs repetitively stacked convolution layers and pooling layers for feature extraction, finalizing the classification through fully connected layers. Multiple 3 × 3 small convolution kernels supplant larger ones, which diminishes the network's parameter count and accelerates convergence. The superimposed

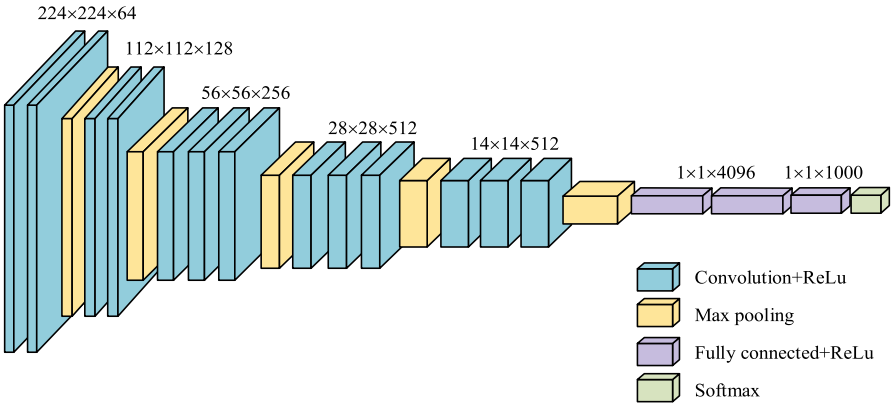


Fig. 1 The network structure of VGG16

convolutional layers augment the network's nonlinear expression capacity and facilitate more white blood cell feature extraction. Small pooling kernels are advantageous for capturing detailed information from white blood cell images.

2.2 ResNet50 network

He et al. proposed the ResNet50 network model in 2015, tackling the issue of network degradation via the introduction of residual learning [16]. It exhibits excellent performance in tasks such as classification, detection, localization, and segmentation. The ResNet50 network structure, displayed in Fig. 2, is partitioned into three components: an input module, four residual modules (each possessing 3, 4, 6, and 3 smaller blocks, respectively), and an output module.

In the ResNet50 network, the input image initially undergoes convolution and pooling operations to produce the $56 \times 56 \times 64$ feature map. Following the computations of the four residual modules, i.e., a sum of 16 smaller blocks, a $7 \times 7 \times 2048$ feature map is procured. Subsequently, an average pooling layer outputs the $1 \times 1 \times 2048$ feature map, and the multi-dimensional data are flattened into a one-dimensional format of 2048 by the Flatten layer. Finally, the sample's category is output by the Softmax function through the fully connected layer.

The notable merit of the ResNet50 network resides in the utilized residual module. By incorporating skip connections between convolutional layers, the output of a preceding layer is directly added to a later layer's output across multiple layers. This structure effectively mitigates the gradient decay issue that transpires with network layer deepening. Consequently, it can better extract high-dimensional features from white blood cell images and yield more accurate classification results.

2.3 Inception V3 network

Szegedy et al. proposed the Inception V3 model [17], a deep convolutional neural network architecture characterized by three kinds of Inception modules, each containing several parallel convolutional layers and pooling layers. Its structure is presented in Fig. 3.

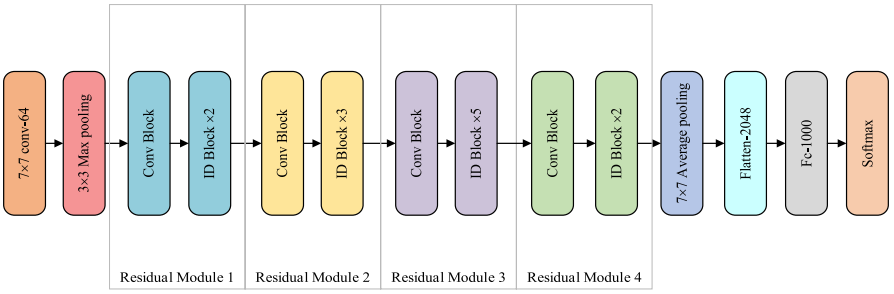


Fig. 2 The network structure of ResNet50

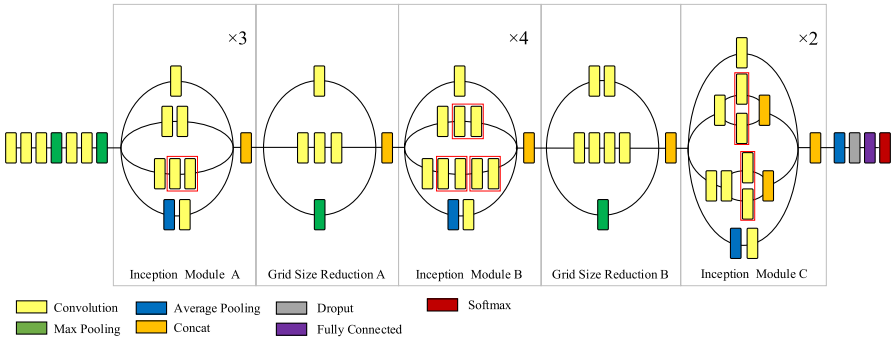


Fig. 3 The network structure of Inception V3

In Inception V3, the Inception module decomposes larger convolutions into smaller ones to reduce parameter quantity. In the Inception A module, the 5×5 convolution is dissected into two 3×3 convolutions. In the Inception B module, the 7×7 convolution is replaced with a 1×7 convolution and a 7×1 convolution. In the Inception C module, the 3×3 convolution is decomposed into a 1×3 convolution and a 3×1 convolution. Conventionally, CNNs utilize pooling operations to downsize the grid size of feature maps. However, in Inception V3, by placing the grid size reduction module between different Inception module types, the computational cost can be further decreased while eliminating the bottleneck.

The Inception V3 network leverages convolution kernels of varied sizes, facilitating receptive fields across different regions. Moreover, the model employs the modular structure to achieve fusion of features at different scales. The Inception module is capable of extracting higher-dimensional features at diverse levels of abstraction, thus allowing for more comprehensive data mining in white blood cell images.

The selection of base models is an important step in ensemble model construction as it directly influences the eventual recognition outcome. It is imperative to choose single models with strong recognition performance, yet maintaining significant distinctions among the base models for information complementarity. In the medical image processing sphere, where labeled samples are scarce and datasets are limited, the selection of not overly complex networks is advantageous for training with restricted white blood cell image data. The superimposed small convolution kernels and pooling kernels in the VGG16 network facilitate more detailed feature extraction. The residual module in the ResNet50 network mitigates gradient vanishing, thereby enabling enhanced deep feature extraction. The Inception V3 network, adopting the Inception modular structure, gleans more diverse and rich information. The distinct structure of these three networks renders them substantially different, hence enabling multi-perspective extraction of different feature information from white blood cell images. Thus, VGG16, ResNet50, and Inception V3 are selected as the base learners.

3 The ensemble method based on the convolutional neural network

The information about white blood cells learned by a singular convolutional neural network is often incomplete, and the accuracy of classification is not necessarily guaranteed by a single convolutional neural network. Ensemble learning combines the learning outcomes of multiple models via the combination strategy, and yields diversified information about white blood cells, thereby bolstering the model's reliability and generalizability.

3.1 Ensemble learning

Ensemble learning integrates multiple base learners through the combination strategy to execute learning tasks like classification and prediction, achieving superior results compared to the single model. The fundamental structure of ensemble learning is depicted in Fig. 4. The dataset is sent to numerous base learners for training, following which the outputs from these learners are combined according to the combination strategy, yielding the final predictive output.

The Bagging ensemble algorithm employs the parallel approach to train multiple base learners simultaneously, integrating their results to procure the final output. There exists no significant dependency between base learners. This study adopts the Bagging method for ensemble with the objective of bolstering the model's generalization and reliability, securing superior predictive outcomes.

3.2 Combination strategy

We adopt the reparameterized Gompertz function to devise a novel combination strategy for ensemble learning. The corresponding fuzzy level is generated by utilizing the prediction confidence of each base learner, and the combination strategy is crafted considering both prediction confidence and fuzzy level. This strategy concurrently accounts for the influence of each base learner's prediction

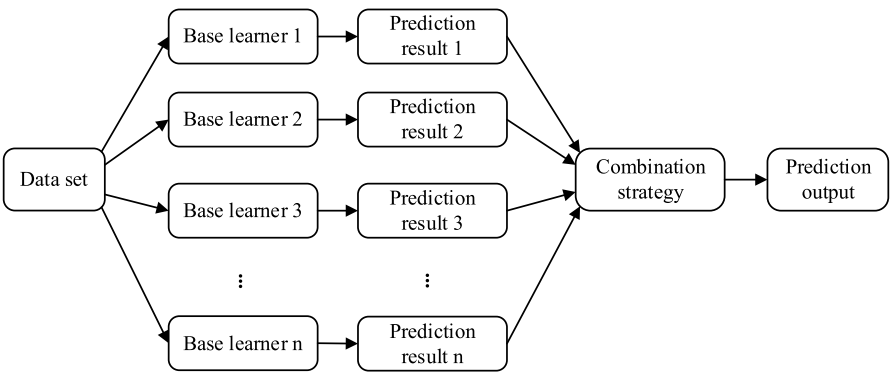


Fig. 4 Schematic diagram of ensemble learning

confidence and fuzzy level on the classification results, allowing for a more comprehensive amalgamation of the prediction results of VGG16, ResNet50, and Inception V3.

The Gompertz function is employed as the animal population growth model to describe population extinction laws [18]. The characteristics of this curve include: initial slow growth reflecting a steady upward trend, mid-term rapid growth illustrating a swift upward trajectory, and eventual slow growth indicating a plateau. The overall evolution of the curve is S-shaped. The Gompertz function's elementary form is illustrated in Eq. (1).

$$y = ae^{-e^{b-ct}} \tag{1}$$

where y represents the research object, t represents time, a, b, and c represent the related parameters.

Since one side of the Gompertz curve invariably approaches the x-axis, it can be inconsistent with certain practical applications. To maximize the value of the Gompertz curve, we modify it by incorporating vertical translation, as shown in Eq. (2).

$$y = 1 - ae^{-e^{b-ct}} \tag{2}$$

The reparameterized Gompertz function is illustrated in Eq. (3).

$$y = 1 - e^{-e^{-2x}} \tag{3}$$

where x represents the base learner's prediction confidence concerning the sample, while y denotes the fuzzy level generated according to the confidence.

Given M classifiers classifying the dataset into N classes, each image $I(x, y)$ will yield $M \times N$ confidences $\left\{ \mathrm{CF}_1^{(1)}, \mathrm{CF}_1^{(2)}, ..., \mathrm{CF}_1^{(N)}, \mathrm{CF}_2^{(1)}, \mathrm{CF}_2^{(2)}, ..., \mathrm{CF}_2^{(N)}, ..., \mathrm{CF}_M^{(1)}, \mathrm{CF}_M^{(2)}, ..., \mathrm{CF}_M^{(N)} \right\}$. In our case, $M = 3$ and $N = 5$, given that three CNN networks are employed. To distinguish five kinds of white blood cells. The confidence is normalized in line with Eq. (4).

$$\sum_{i=1}^{M} \sum_{j=1}^{N} \mathrm{CF}_i^{(j)} = 1 \tag{4}$$

Where $\forall i = 1, 2, ..., M; \forall j = 1, 2, ..., N$, M represents the number of classifiers, and N represents the number of classes in the dataset.

The re-parameterized Gompertz function is harnessed to generate the fuzzy level corresponding to each confidence, the computation of which is represented in the following equation.

$$\mathrm{FL}_i^{(j)} = 1 - \exp\left(-\exp\left(-2 \times \mathrm{CF}_i^{(j)}\right)\right) \tag{5}$$

The confidence value lies between 0 and 1. When $CF_i^{(j)}$=1, $FL_i^{(j)}$ obtains the minimum value of 0.127, while when $CF_i^{(j)}$=0, $FR_i^{(j)}$ obtains reaches the maximum value of 0.632. Consequently, the value range of $FL_i^{(j)}$ spans from 0.127 to 0.632. The higher the confidence, the lower the fuzzy level. The mean confidence complement and the fuzzy level sum are calculated via Eqs. (6) and (7), respectively.

$$\mathrm{CMCF}^{(j)} = 1 - \frac{1}{M}\sum_{i=1}^{M} \mathrm{CF}_i^{(j)} \tag{6}$$

$$\mathrm{SFL}^{(j)} = \sum_{i=1}^{M} \mathrm{FL}_i^{(j)} \tag{7}$$

The final decision score is obtained by multiplying the complement of mean confidence and the sum of fuzzy level, as computed by the following equation.

$$\mathrm{DS}^{(j)} = \mathrm{CMCF}^{(j)} \times \mathrm{SFL}^{(j)} \tag{8}$$

The sample's category is ascertained based on the minimal decision score, as shown in Eq. (9).

$$\mathrm{Class}(I) = \arg\min \left\{ \mathrm{DS}^{(1)}, \mathrm{DS}^{(2)}, \ldots, \mathrm{DS}^{(N)} \right\} \tag{9}$$

Correct predictions can be made via the calculation of the decision score of the samples and identifying the category corresponding to the minimal decision score.

3.3 The proposed method

This study selects three distinct convolutional neural networks with excellent performance and diverse structures as base learners. Subsequently, the prediction results of these base learners are combined through the formulated combination strategy. By utilizing the characteristics of various network models and integrating the diversity information of white blood cell images, an ensemble learning model with superior performance is achieved, thereby enhancing the overall performance. The framework of the proposed method is delineated in Fig. 5.

As observable in Fig. 5, the white blood cell dataset is initially sent to the VGG16, ResNet50, and Inception V3, respectively. These three convolutional neural networks, based on transfer learning, classify white blood cell samples, presenting the prediction results as probabilities. The fuzzy level of each base learner is generated based on the improved Gompertz function, and the final category prediction for the white blood cells is determined by calculating the ultimate decision score. The flowchart of the proposed algorithm is portrayed in Fig. 6.

Classification necessitates initial preprocessing of the white blood cell dataset. The training set and the validation set are partitioned in a 4:1 ratio. Subsequently, three pre-trained CNN models are fine-tuned on the white blood cell dataset. The

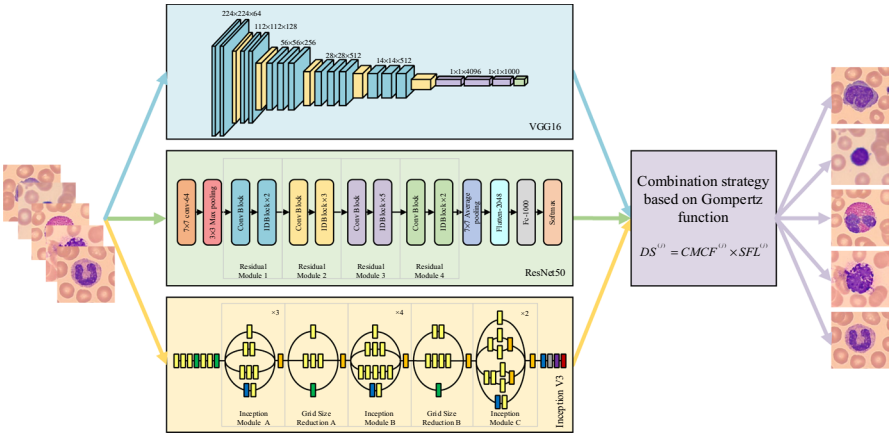


Fig. 5 Framework of the proposed method

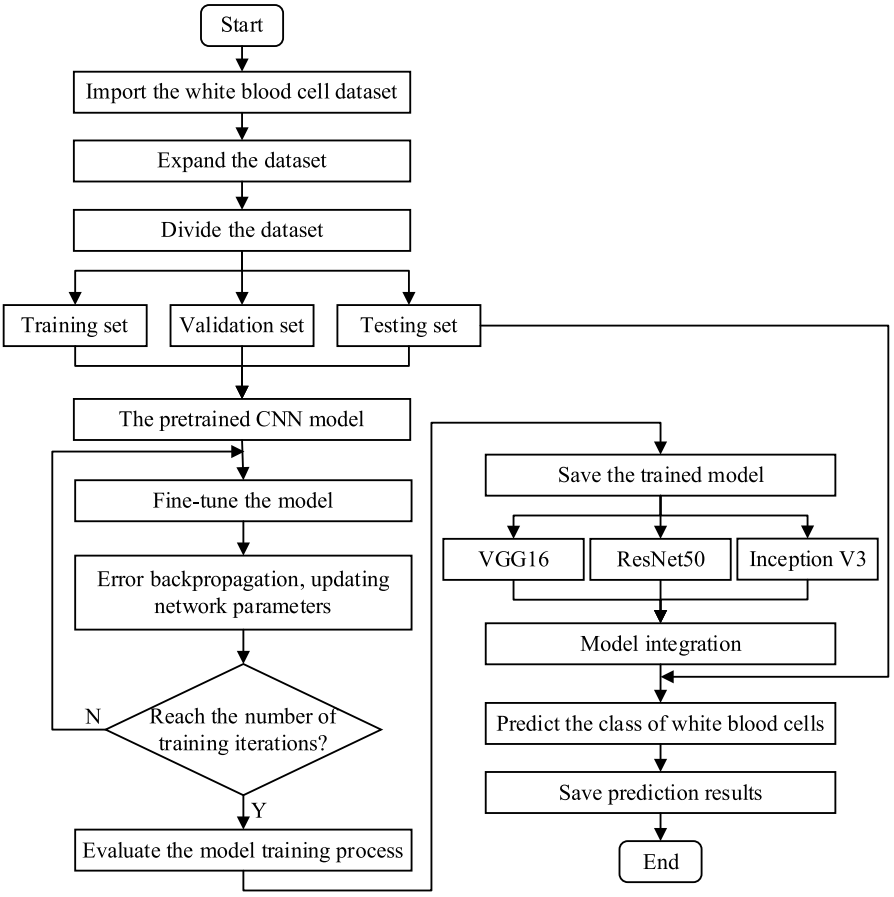


Fig. 6 Flowchart of the proposed method

iterative process halts once the prediction accuracy requirement is met or upon reaching maximum iteration. The parameters and prediction results of each network model are stored. Lastly, the learning results of the VGG16, ResNet50, and Inception V3 are integrated via Bagging, and the combination strategy is formulated based on the improved Gompertz function. By integrating the prediction results of the base models, the category corresponding to the minimal decision score is utilized as the ultimate prediction result.

4 Experimental results and analysis

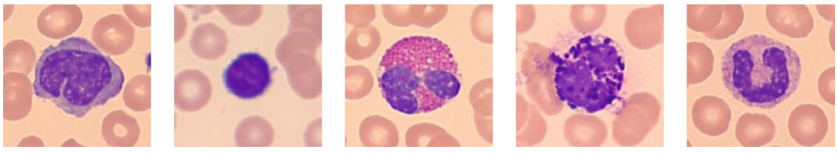
4.1 Dataset and experiment setup

A substantial number of blood cell smears are collated from a public dataset (https://github.com/wangqw062/leukocyteData). Image enhancement technology [19] is used to augment the images and form our exclusive white blood cell dataset. Data augmentation elevates the picture count by threefold, effectively circumventing model over-fitting during training. This dataset comprises five types of white blood cells: monocytes, lymphocytes, eosinophils, basophils, and neutrophils, as displayed in Fig. 7.

The experiment is executed on the Pytorch platform, with each single model pre-trained on the ImageNet dataset. To closely monitor the results of model training, the iteration of each model is set to 50, the batch size to 32, the learning rate to 0.001, and the momentum to 0.9. Furthermore, the input image size for the VGG16 and ResNet50 networks is 224×224, while for the Inception V3 network, it is 299×299. The training set contains 320 samples of each type of white blood cell, while the validation set includes 80.

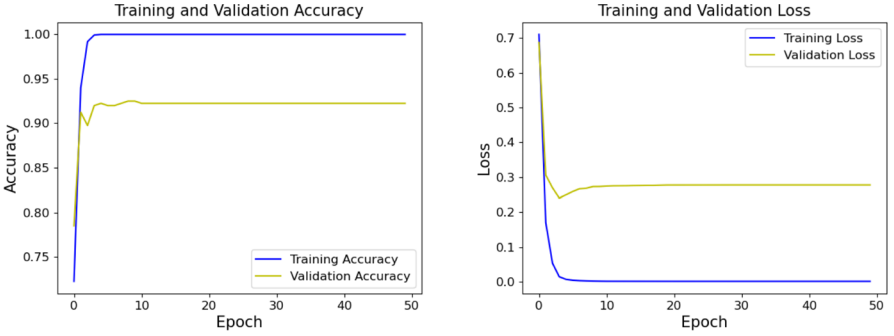
4.2 Learning results of the single model

This paper employs three distinct CNN models, namely VGG16, ResNet50, and Inception V3, as base learners. These models, with their varying structures and depths, are pre-trained on the ImageNet dataset, then transferred for fine-tuning on the white blood cell dataset. The accuracy and loss rate variation curves of each base learner during the iterative process are depicted in Fig. 8.



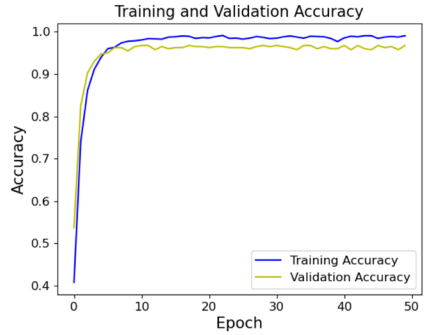
(a) monocyte (b) lymphocyte (c) eosinophil (d) basophil (e) neutrophil

Fig. 7 Diagram of white blood cells

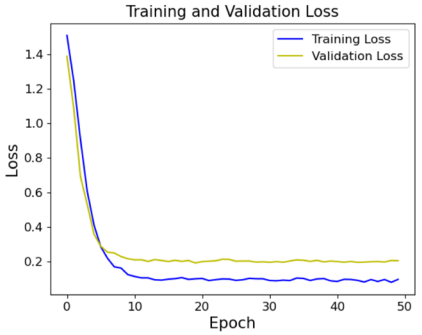


(a) The accuracy of VGG16

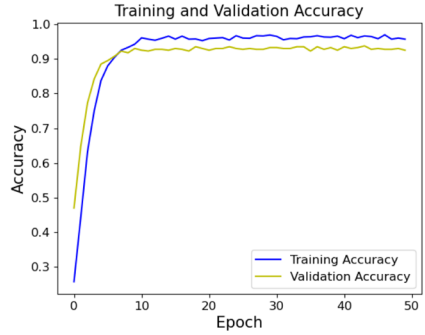
(b) The loss of VGG16



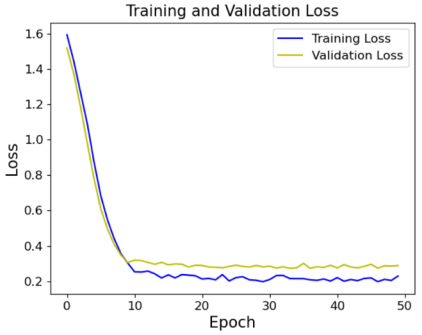
(c) The accuracy of ResNet50



(d) The loss of ResNet50



(e)The accuracy of Inception V3



(f) The loss of Inception V3

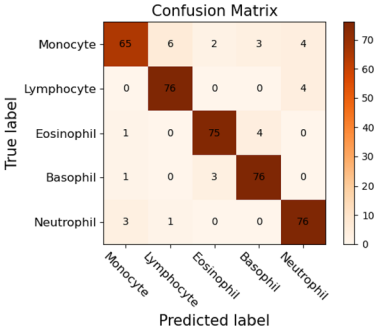
Fig. 8 The curve of accuracy and loss rate of the base learner

As demonstrated in Fig. 8, the accuracy of each base learner on both the training and validation sets continually increases with the number of iterations, while the loss rate exhibits a gradual decrease. Approximately at 30 epochs, the accuracy and loss rates for the VGG16, ResNet50, and Inception V3 on the training and validation

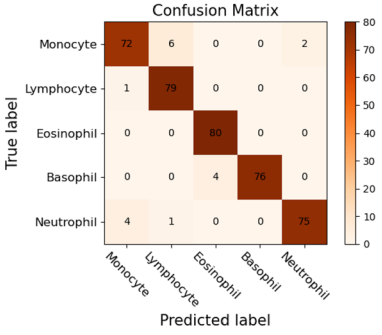
sets stabilize, indicating competent performance of each base learner on the white blood cell dataset.

The confusion matrix, an evaluative metric for assessing model classification performance, expresses the correlation between samples' actual attributes and predicted outcomes in a matrix format. The matrix rows represent the true values, and the columns signify the predicted values. More specifically, the total data in each row of the confusion matrix represents the true number of certain samples; the total data in each column corresponds to the predicted number of certain samples, and the diagonal value indicates the correct prediction quantity for certain samples. This paper employs a 5×5 matrix to present the classification outcomes of five types of white blood cells. The classification results for each base learner are exhibited in Fig. 9.

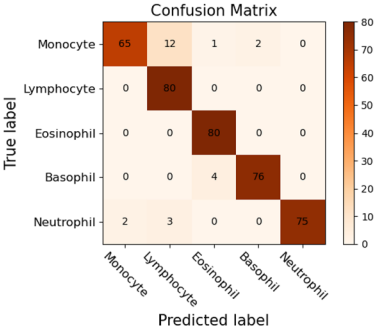
In the classification of 400 white blood cell images, VGG16, ResNet50, and Inception V3 yield average recognition accuracies of 92%, 95.5%, and 94% respectively. If the performance of a given base model is subpar compared to the others, the model could potentially introduce noise and negatively impact on the final recognition accuracy. However, as VGG16, ResNet50, and Inception V3 exhibit consistent recognition of different types of white blood cells, minimal



(a) Classification results of VGG16



(b) Classification results of ResNet50



(c) Classification results of InceptionV3

Fig. 9 Classification results of each base learner

Table 1 Classification performance of each base learner

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
VGG16	92.00	92.04	92.00	91.92
ResNet50	95.50	95.60	95.50	95.49
Inception V3	94.00	94.04	94.00	93.72

disparity exists in the classification effects among base learners, thereby avoiding the aforementioned problem.

The classification performance of each base learner is further evaluated in this paper using four assessment indices: accuracy, precision, recall, and F1-score, as presented in Table 1. The accuracy, precision, recall, and F1-score of VGG16, ResNet50, and Inception V3 all exceed 90%, indicating their robust classification capabilities.

4.3 Evaluation of the ensemble model

In this study, we employ the devised combination strategy to integrate VGG16, ResNet50, and InceptionV3. The final category of white blood cells is achieved through the constructed ensemble learning model. The classification results yielded by this ensemble learning model are demonstrated in Fig. 10.

As is seen from Fig. 10, the proposed methodology accurately classifies 386 out of 400 images. From these accurately classified images, 71 monocytes, 80 lymphocytes, 80 eosinophils, 77 basophils, and 78 neutrophils. Misclassifications include

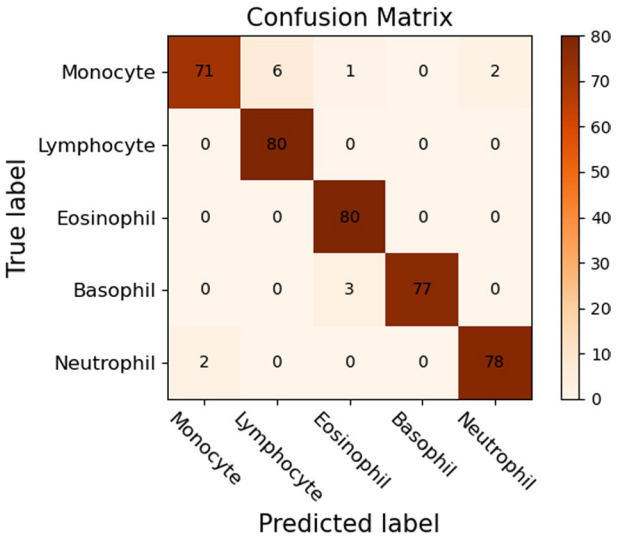


Fig. 10 Classification results of the ensemble learning model

9 monocytes as other cell types, 3 basophils as eosinophils, and 2 neutrophils as monocytes. A potential explanation for these misclassified cells could be insufficient distinctive information in the images. Despite a small proportion of cells being misclassified, the number of such samples is sufficiently low, offering promising potential for auxiliary diagnostic applications.

Sensitivity or True Positive Rate (TPR) denotes the percentage of actual positive samples predicted as positive. A higher value, ideally 1, signifies better. Specificity or False Positive Rate (FPR) is the proportion of actual negative samples predicted as positive. The smaller this value, ideally 0, the better. The Receiver Operating Characteristic (ROC) curve is plotted with TPR on the y-axis and FPR on the x-axis, with an ideal point at (0, 1), where TPR = 1 and FPR = 0. Hence, the closer the ROC curve approaches the upper left corner, the superior the model's classification performance and prediction accuracy. The Area Under Curve (AUC) reflects the model's classification performance, typically ranging between 0.5 and 1. A larger AUC signifies superior classification effects. When the AUC is closer to 1, prediction is more accurate; when AUC falls between 0.5 and 1, the classifier has predictive value. AUC = 0.5 indicates no predictive value. The ROC curve obtained via the ensemble learning model is depicted in Fig. 11.

The blue solid line, yellow dotted line, green point line, red dotted line, and purple point line represent the ROC curves of monocytes, lymphocytes, eosinophils, basophils, and neutrophils, respectively. It is evident from Fig. 11 that the green point line lies closer to the axis's upper left corner, demonstrating the classifier's superior ability to distinguish eosinophils. Conversely, the blue solid line is the furthest from the upper left corner, indicating the classifier's lower identification performance for monocytes. The areas under both the yellow dotted line and the green point line are 0.99, indicating an AUC of 0.99 for both lymphocytes and eosinophils with the proposed method, achieving superior classification performance for these two types of cells. The area under the blue solid line is 0.94, indicating that the AUC

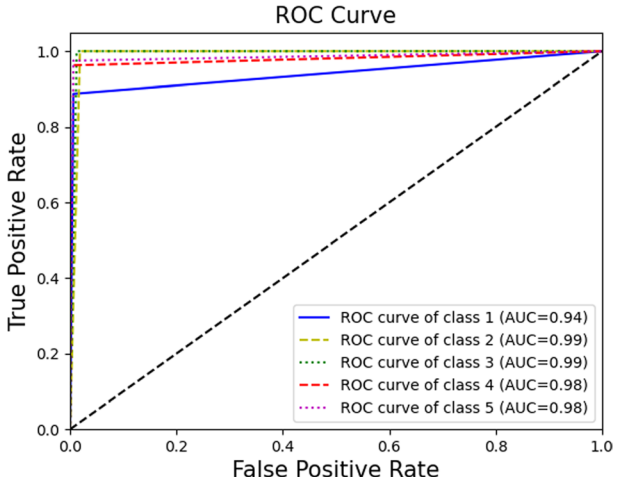


Fig. 11 ROC curve of the ensemble learning model

Table 2 Classification performance of the ensemble learning model for different types of cells

Type	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Monocyte	88.75	97.26	88.75	92.81
Lymphocyte	100.00	93.02	100.00	96.39
Eosinophil	100.00	95.24	100.00	97.56
Basophil	96.25	100.00	96.25	98.09
Neutrophil	97.50	97.50	97.50	97.50
Average	96.50	96.60	96.50	96.47

of the proposed method for monocytes is 0.94, implying lower classification accuracy for this type of cells.

This paper further verifies the constructed ensemble learning model's classification performance by employing four evaluation metrics: accuracy, precision, recall, and F1-score, as displayed in Table 2.

Upon recognition of 400 white blood cell images, a mere 9 monocytes, 3 basophils, and 2 neutrophils are misclassified, culminating in an overall classification accuracy of 96.5%. As delineated in Table 2, the recognition accuracy of lymphocytes and eosinophils utilizing the proposed method is 100%, with precisions of 93.02% and 95.24%, respectively. Both types present recalls of 100%, and their F1-scores are 96.39% and 97.56%, respectively. The recognition accuracy for basophils and neutrophils is 97.5% and 96.25%, respectively, with precisions of 97.5% and 100%, respectively. The recalls for these two types are 97.5% and 96.25%, respectively, with F1-scores of 97.5% and 98.09%, respectively. The recognition accuracy of monocytes, though not ideal at 88.75%, still reflects a commendable performance level.

The obtained performance indicators are visually represented via a bar graph as displayed in Fig. 12. The proposed method demonstrates superior results for

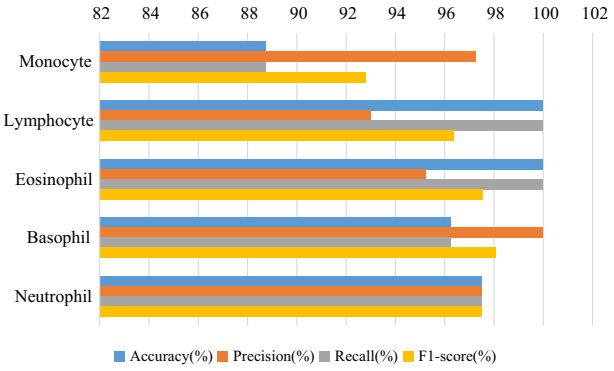


Fig. 12 Classification performance of the ensemble learning model for different types of cells

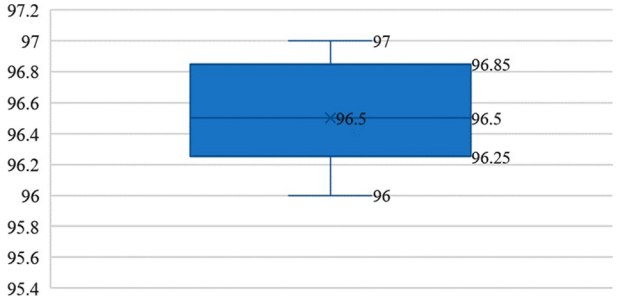


Fig. 13 Recognition accuracy of the ensemble learning model

lymphocytes, eosinophils, basophils, and neutrophils, while the recognition of monocytes is less effective, potentially due to indistinguishable features in some of the images.

This study employs a ten-fold cross-validation for training and testing, and evaluates the algorithm based on the average of ten classification outcomes, which are presented as a box diagram in Fig. 13. The average accuracy of the proposed method is 96.5%, with a fluctuation within 0.5%, affirming the efficacy of the constructed ensemble learning model. Ensemble learning can improve the reliability of white blood cell classification and effectively reduce the fluctuation of classification results.

To further underscore the benefits of the ensemble learning model, we compare the single architecture of VGG16, ResNet50, and InceptionV3 to the ensemble learning model, as demonstrated in Fig. 14.

These three individual models with obvious structural differences, all achieve a recognition accuracy surpassing 90%. The formulated combination strategy integrates VGG16, ResNet50, and InceptionV3. As observed in Fig. 14, the

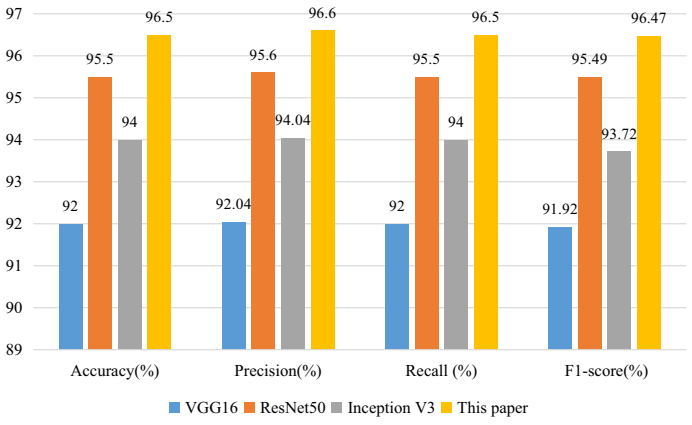


Fig. 14 Comparison of classification performance of different models

Table 3 Experimental results of integration based on average method

	Monocyte	Lymphocyte	Eosinophil	Basophil	Neutrophil
Monocyte	71	6	1	0	2
Lymphocyte	0	80	0	0	0
Eosinophil	0	0	80	0	0
Basophil	0	0	3	77	0
Neutrophil	4	0	0	0	76

Table 4 Experimental results of integration based on voting method

	Monocyte	Lymphocyte	Eosinophil	Basophil	Neutrophil	Unable to judge
Monocyte	72	5	0	0	1	2
Lymphocyte	0	80	0	0	0	0
Eosinophil	0	0	80	0	0	0
Basophil	0	0	3	75	0	2
Neutrophil	2	0	0	0	76	2

Table 5 Comparison of the accuracy of integration using different combination strategies

Method	Monocyte	Lymphocyte	Eosinophil	Basophil	Neutrophil	Average accuracy (%)
The average method	88.75	100.00	100.00	96.25	95.00	96.00
The voting method	90.00	100.00	100.00	93.75	95.00	95.75
Our proposed method	88.75	100.00	100.00	96.25	97.50	96.50

classification performance of the ensemble learning model surpasses that of any base learner, attesting to the superiority of the proposed ensemble learning model. It unequivocally shows that ensemble learning can enhance the classification efficacy of white blood cells.

Additionally, the study applies averaging and voting methods to combine base learners, the results of which are detailed in Tables 3 and 4. From Table 3, it is evident that the averaging method fails to correctly identify 16 white blood cells, including 9 monocytes, 3 eosinophils, and 4 neutrophils, with an average recognition accuracy of 96%. Table 4 reveals that the voting method results in 17 white blood cells not being correctly identified, with 11 incorrectly recognized and 6 not directly identified, with an average recognition accuracy of 95.75%.

The diversity of models can augment the recognition capability of white blood cell images, while different combination strategies can exert certain impact on the recognition efficacy of ensemble learning models. To authenticate the preeminence of the proposed combination strategy, the same dataset and identical base learners

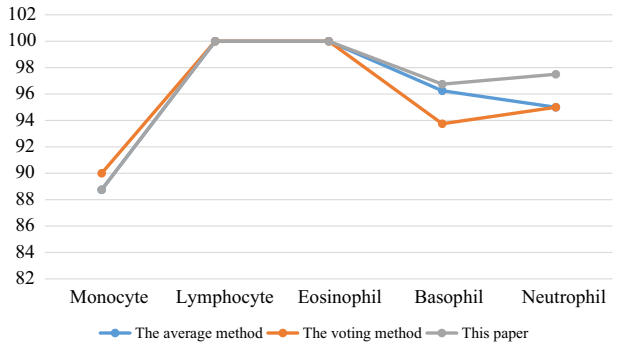


Fig. 15 Comparison of different combination strategies

Table 6 Comparison of classification performance between the proposed method and other methods

Reference	Method	Sub method	Acc (%)	Pr (%)	Re (%)	F1 (%)
Banik et al. [20]	Fused CNN	–	95.19	92.32	90.39	90.62
Patil et al. [21]	CNN + RNN	VGG16 + LSTM	89.95	90.34	88.86	93.19
		InceptionV3 + LSTM	91.66	91.03	90.96	90.43
		ResNet50 + LSTM	93.16	94.70	93.09	93.09
Cengil et al. [22]	Fused CNN + ML	AlexNet + ResNet + GoogleNet & LDA	93.60	93.55	93.53	93.48
		AlexNet + ResNet + GoogleNet & SVM	92.20	92.18	92.10	92.08
Proposed method	Ensembled CNN	VGG16 + ResNet50 + Inception V3	**96.50**	**96.60**	**96.50**	**96.47**

(VGG16, ResNet50, and InceptionV3) are employed, with the averaging and voting methods serving as comparative measures, as demonstrated in Table 5.

As Table 5 reveals, despite the proposed combination strategy's slightly inferior recognition accuracy for monocytes compared to the voting method, its average recognition accuracy surpasses both the voting and averaging methods. This proves the proposed combination strategy's superiority, which optimally integrates each base learner and leverages the advantages of disparate models.

For a more intuitive comparison of the experimental outcomes of different combination strategies, a line chart is illustrated in Fig. 15. Relative to the averaging method, the proposed combination strategy's recognition accuracy for monocytes, lymphocytes, eosinophils, and basophils aligns, while the accuracy for neutrophils is markedly higher. In comparison with the voting method, despite the proposed combination strategy's marginally lower recognition accuracy for monocytes, it nonetheless remains above 88%, with the accuracy for eosinophils and neutrophils considerably higher.

To further verify the proposed method's efficacy, Table 6 offers a performance comparison between the proposed method and certain existing techniques. Banik et al. [20] introduced a fused CNN model where the feature maps of two

convolutional layers were integrated via max pooling operation, then input to the fully-connected neural network layer to classify white blood cell images. Patil et al. [21] proposed a fusion model of CNN and RNN with canonical correlation analysis for blood cell image classification, achieving superior classification accuracy by enhancing the comprehension of the image content. Cengil et al. [22] initially extracted features from certain layers of three pre-trained CNN networks and combined these features. Subsequently, they employed Principal Component Analysis for feature reduction and selected traditional machine learning classifiers for white blood cell classification.

As shown in Table 6, when classifying white blood cells, the accuracy, precision, recall, and F1-score attained by the proposed ensemble learning model exceed 96%, which are the highest and represented in bold. The comparison method's average accuracy is 92.63%, with the minimum and maximum accuracy being 89.95% and 95.19% respectively. The proposed method's average accuracy is 96.5%, outperforming the comparison method's average accuracy by 3.87%, exceeding the minimum accuracy by 6.55% and surpassing the maximum accuracy by 1.31%. Furthermore, the proposed method excels over the comparison method in precision, recall, and F1-score, thereby substantiating its robust classification capability.

5 Summary and outlook

Exploiting the automated learning capabilities of convolutional neural networks enables the effective circumvention of the intricacies involved in cell segmentation and disparities arising from manually selected features. Notably, when classifying white blood cells, the information derived from a single network tends to be somewhat unilateral, and a richer spectrum of information can be harvested through the integration of multiple networks. Accordingly, this study introduces a white blood cell classification methodology based on convolutional neural networks and ensemble learning. Primarily, three disparate convolutional neural networks are selected as base learners, followed by the combination of these base learners' results, premised on the devised combination strategy, to ascertain the ultimate category. By combining various network models, the proposed method learns diverse feature information of white blood cells, thereby markedly enhancing the classification's reliability. With an accuracy rate of 96.5% on our white blood cell dataset, this method surpasses the compared ones, thus conclusively manifesting its superiority.

The current experimental procedures are conducted on single-cell images. Future research will take into account the influence of overlapping nucleus and cell clusters on the classification outcomes. Additionally, other types of convolutional neural networks and different combination strategies can be tried to improve the classification performance of the model.

Acknowledgments This work was funded by the National Natural Science Foundation of China under Grant 62273253.

Author's contribution ND contributed to funding acquisition, supervision, writing—review and editing. QF contributed to methodology, writing—original draft, review and editing. JC contributed to formal analysis and investigation. XM contributed to data collection and analysis.

Data availability The datasets used during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Li Y, Zhu R, Mi L, Cao YH, Yao D (2016) Segmentation of white blood cell from acute lymphoblastic leukemia images using dual-threshold method. Comput Math Methods Med 2016:9514707. https://doi.org/10.1155/2016/9514707
2. Zheng X, Wang Y, Wang GY, Liu JG (2018) Fast and robust segmentation of white blood cell images by self-supervised learning. Micron 107:55–71. https://doi.org/10.1016/j.micron.2018.01.010
3. Pergad ND, Hamde ST (2018) Fractional gravitational search-radial basis neural network for bone marrow white blood cell classification. Imaging Sci J 66(2):106–124. https://doi.org/10.1080/13682199.2017.1383677
4. Tavakoli S, Ghaffari A, Kouzehkanan ZM, Hosseini R (2021) New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images. Sci Rep 11(1):19428. https://doi.org/10.1038/s41598-021-98599-0
5. Khan MA, Qasim M, Lodhi HMJ, Nazir M, Javed K, Saddaf R, Din A, Habib U (2020) Automated design for recognition of blood cells diseases from hematopathology using classical features selection and ELM. Microsc Res Tech 84(2):202–216. https://doi.org/10.1002/jemt.23578
6. Ahmed L, Iqbal MM, Aldabbas H, Khalid S, Saleem Y, Saeed S (2020) Images data practices for semantic segmentation of breast cancer using deep neural network. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-01680-1
7. Kumari S, Bhatia M (2022) A cognitive framework based on deep neural network for classification of coronavirus disease. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-022-03756-6
8. Kumar A, Kalia A, Sharma A, Kaushal M (2021) A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-021-03541-x
9. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://doi.org/10.1145/3065386
10. Jiang M, Cheng L, Qin FW, Du L, Zhang M (2018) White blood cells classification with deep convolutional neural networks. Int J Pattern Recognit Artif Intell 32(9):1857006. https://doi.org/10.1142/S0218001418570069
11. Banik PP, Saha R, Kim KD (2020) An automatic nucleus segmentation and CNN model based classification method of white blood cell. Expert Syst Appl 149:113211. https://doi.org/10.1016/j.eswa.2020.113211
12. Bagido RA, Alzahrani M, Arif M (2021) White Blood cell types classification using deep learning models. Int J Comput Sci Netw Secur 21(9):223–229. https://doi.org/10.22937/IJCSNS.2021.21.9.30
13. Liang GB, Hong HC, Xie WF, Zheng LX (2018) Combining convolutional neural network with recursive neural network for blood cell image classification. IEEE Access 6:36188–36197. https://doi.org/10.1109/ACCESS.2018.2846685

14. Mohamed EH, El-Behaidy WH, Khoriba G, Li J (2020) Improved white blood cells classification based on pre-trained deep learning models. J Commun Softw Syst 16(1):37–45
15. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 2015 International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1409.1556
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp770–778. https://doi.org/10.1109/CVPR.2016.90
17. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2818–2826. https://doi.org/10.1109/CVPR.2016.308
18. Gompertz B (1825) On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philos Trans R Soc Lond 115:513–585. https://doi.org/10.1098/rstl.1825.0026
19. Huang L, Pan WJ, Zhang Y, Qian LP, Gao N, Wu Y (2020) Data augmentation for deep learning-based radio modulation classification. IEEE Access 8:1498–1506. https://doi.org/10.1109/ACCESS.2019.2960775
20. Banik PP, Saha R, Kim KD (2019) Fused convolutional neural network for white blood cell image classification. In: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp 238–240. https://doi.org/10.1109/ICAIIC.2019.8669049
21. Patil AM, Patil MD, Birajdar GK (2021) White blood cells image classification using deep learning with canonical correlation analysis. IRBM 42(5):378–389. https://doi.org/10.1016/j.irbm.2020.08.005
22. Cengil E, Cinar A, Yildirim M (2021) A hybrid approach for efficient multi-classification of white blood cells based on transfer learning techniques and traditional machine learning methods. Concurr Comput-Pract Exp 34(6):e6756. https://doi.org/10.1002/cpe.6756

Authors and Affiliations

Na Dong[1] · Qingyue Feng[1] · Jianfang Chang[1] · Xiaoming Mai[1]

✉ Na Dong
dongna@tju.edu.cn

Qingyue Feng
2437739689@qq.com

Jianfang Chang
changjianfang@tju.edu.cn

Xiaoming Mai
13903000924@126.com

[1] School of Electrical and Information Engineering, Tianjin University, 92 Weijin Road, Nankai District, Tianjin, China