Indonesian Journal of Data and Science

Volume 6 Issue 2
ISSN 2715-9936
https://doi.org/10.56705/ijodas.v6i2.300

Research Article

Deep Learning-Based Blood Cell Image Classification Using ResNet18 Architecture

Thomas Edyson Tarigan [1,*]; Agung Budi Prasetyo [2]; Emy Susanti [3]
[1] Universitas Teknologi Digital Indonesia, Yogyakarta, 55198, Indonesia, tarigan@utdi.ac.id
[2] Universitas Teknologi Digital Indonesia, Yogyakarta, 55198, Indonesia, Indonesia, agun_bp@utdi.ac.id
[3] Universitas Teknologi Digital Indonesia, Yogyakarta, 55198, Indonesia 9, Indonesia, emysusanti@utdi.ac.id

Correspondence should be addressed to Thomas Edyson Tarigan; tarigan@utdi.ac.id

Received 10 March 2025; Accepted 15 June 2025; Published 31 July 2025

Abstract:
The classification of white blood cells (WBC) plays a critical role in haematological diagnostics, yet manual examination remains a labour-intensive and subjective process. In response to this challenge, this study investigates the application of deep learning, specifically the ResNet18 convolutional neural network architecture, for the automated classification of blood cell images into four classes: eosinophils, lymphocytes, monocytes, and neutrophils. The dataset used comprises microscopic images annotated by cell type and is divided into training and validation sets with an 80:20 ratio. Standard pre-processing techniques such as image normalization and augmentation were applied to enhance model robustness and generalization. The model was fine-tuned using transfer learning with pre-trained weights from ImageNet and optimized using the Adam optimizer. Performance was evaluated through a comprehensive set of metrics including accuracy, precision, recall, F1-score, mean squared error (MSE), and root mean squared error (RMSE). The best model achieved a validation accuracy of 86.89%, with macro-averaged precision, recall, and F1-score of 0.8738, 0.8690, and 0.8688, respectively. Lymphocyte classification yielded the highest F1-score (0.9515), while eosinophils posed the greatest classification challenge, as evidenced by lower precision and higher misclassification rates in the confusion matrix. Error-based evaluation further supported the model's consistency, with an MSE of 0.7125 and RMSE of 0.8441. These results confirm that ResNet18 is capable of learning discriminative features in complex haematological imagery, providing an efficient and reliable alternative to manual analysis. The findings suggest potential for practical implementation in clinical workflows and pave the way for further research involving multi-model ensembles or cell segmentation pre-processing for improved precision.

Keywords: Blood Cell Classification, ResNet18, Deep Learning, Medical Imaging, Multi-class Evaluation.

Dataset link: https://www.kaggle.com/datasets/paultimothymooney/blood-cells

1. Introduction

The microscopic landscape of human blood harbors a wealth of diagnostic information—yet its interpretation remains largely manual, constrained by human fatigue, and prone to inconsistency. While hematology has benefited from laboratory automation, the classification of white blood cells (WBC) into specific types such as eosinophils, lymphocytes, monocytes, and neutrophils are still dominated by expert visual inspection under a microscope. This reliance on subjective assessment introduces not only variability but also a bottleneck in clinical decision-making, especially in high-throughput environments.

Recent advances in deep learning have redefined what is computationally feasible in image classification tasks. Architectures such as ResNet18 offer a promising framework for capturing morphological nuances in complex biological data. Yet, a recurring limitation in existing literature is the tendency to either focus on binary classification problems or to evaluate model performance using a narrow set of metrics—often restricted to accuracy alone. Such

an approach masks underlying weaknesses, particularly in class-imbalanced settings or when false negatives carry significant clinical consequences.

Current studies reflect an expanding interest in adapting deep learning models to medical and biological image classification. [1] combined deep neural networks with particle swarm optimization for radiological analysis, demonstrating how multi-stage learning can suppress false positives. [2] working in the agricultural domain, revealed that ResNet can effectively distinguish between highly similar visual patterns—analogous to the morphological overlap among WBC subtypes. [3] employed morphological features for drug-resistance prediction in cancer cells, indicating the growing trend of non-invasive, image-driven diagnostics. In parallel, [4] explored MobileNetV2 for pulmonary cancer classification via transfer learning, presenting an efficiency-focused alternative to deeper architectures. [5] proposed a hybrid U-Net and CNN pipeline to classify 3D cellular structures, underlining the importance of pre-segmentation in challenging imaging contexts.

These studies establish a technological baseline but stop short of offering a full-spectrum evaluation of deep models on leukocyte classification—an area where both model architecture and metric selection matter. The present research responds to this gap by systematically applying the ResNet18 architecture to classify blood cell images into four morphologically distinct WBC categories. Unlike prior works, we conduct a multidimensional performance evaluation using not only accuracy, but also precision, recall, F1-score, mean squared error (MSE), and root mean squared error (RMSE). By doing so, we aim to clarify both the strengths and potential pitfalls of using ResNet18 in haematological imaging, offering new insights into its clinical applicability.

2. Method

This study adopts an experimental design cantered on the application of a convolutional neural network (CNN), specifically ResNet18 [6]–[8], to the multi-class classification of peripheral blood cell images. The methodology is organized into five core components: dataset acquisition and preparation, model configuration, training procedures, evaluation metrics, and analytical interpretation.

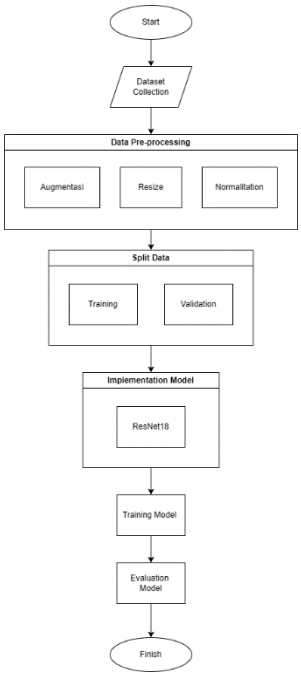


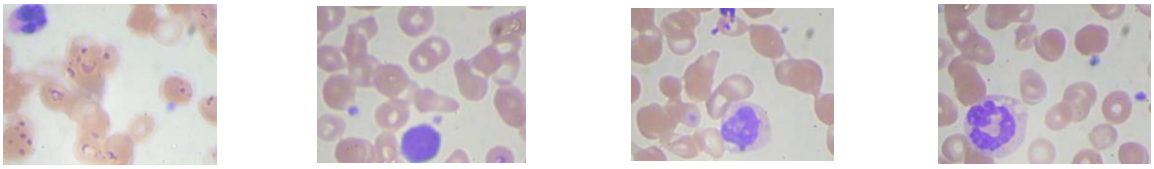
Figure 1: Research Workflow

Dataset and Pre-processing:

We utilize the publicly available Blood Cell Images dataset, which comprises labelled microscopic images of four primary leukocyte classes: eosinophils, lymphocytes, monocytes, and neutrophils. Each class represents distinct morphological patterns, yet inter-class visual similarities pose a non-trivial classification challenge. The dataset is

divided into training, validation, and test sets following an 80:10:10 split ratio to ensure generalization while minimizing overfitting.

Prior to training, all images are resized to 224×224 pixels to match ResNet18's input specification [9]–[11]. Standard image normalization and augmentation techniques—such as random rotation, horizontal flipping, and contrast adjustment—are applied to enhance dataset diversity and mitigate class imbalance effects [12]–[14]. These transformations aim to simulate natural variability in blood smear imaging conditions.



Eosinophil Lymphocyte Monocyte Neutrophil

Figure 2. Sample Images from the Dataset

Model Architecture:

ResNet18, a residual learning-based CNN with 18 layers, is chosen for its architectural efficiency and proven performance in biomedical image classification [15], [16]. Its key innovation—skip connections—enables stable training of deeper networks without gradient degradation [1], [17]. The model is initialized with pre-trained weights from ImageNet to leverage transfer learning [18], then fine-tuned on the blood cell dataset. The final fully connected layer is modified to output four logits corresponding to the target classes.

Training Procedure

Training is conducted using the PyTorch framework. We employ the Adam optimizer with a learning rate of 0.0001 and a batch size of 32. The model is trained over 50 epochs with early stopping based on validation loss to prevent overfitting. Cross-entropy loss is used as the primary objective function due to the categorical nature of the classification task.

Evaluation Metrics

To comprehensively assess the classification performance of the ResNet18 model, we employ a suite of evaluation metrics that capture both categorical accuracy and numerical prediction deviation [19], [20]. These metrics include accuracy, precision, recall, F1-score, mean squared error (MSE), and root mean squared error (RMSE) [21]–[23].

Let the four classes be denoted as $C = \{c_1, c_2, c_3, c_4\}$, corresponding to eosinophils, lymphocytes, monocytes, and neutrophils. Let TP_i, FP_i, FN_i, and TN_i denote the true positives, false positives, false negatives, and true negatives for class c_i, respectively:

a. Accuracy is defined as the ratio of correctly classified samples to the total number of samples:

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

b. Precision for each class measures the proportion of true positives among all positive predictions:

$$Precision_i = \frac{TP_i}{TP_i + FP_I} \tag{2}$$

c. Recall (or sensitivity) evaluates the proportion of actual positives correctly identified:

$$Recall_i = \frac{TP_i}{TP_i + FN_I} \tag{3}$$

d. F1-score is the harmonic mean of precision and recall:

$$F1_i = 2\frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{4}$$

To generalize across classes, we report the macro-averaged values of precision, recall, and F1-score:

$$Macro - X = \frac{1}{|C|}\sum_{i=1}^{|C|} X_i \ where\ X \in \{Precision, Recall, F1\} \quad (5)$$

e. Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) treat the class labels as ordinal indices and measure deviation between predicted class index $\hat{y}$ and true class index y [24], [25]:

$$MSE = \frac{1}{N}\sum_{j=1}^{n}\left(\hat{y}_j - y_j\right)^2 \quad (6)$$
$$RMSE = \sqrt{MSE}$$

This multi-metric evaluation strategy is designed to uncover subtle performance differences—particularly in imbalanced or morphologically overlapping classes—where accuracy alone may be misleading.

3. Result and Discussion

Results

The training process was executed over 10 epochs, during which the model showed a consistent improvement in classification performance. The training accuracy increased from 90.72% in the first epoch to 99.55% in the final epoch, while validation accuracy peaked at 86.89%, indicating that the model was able to generalize well on unseen data. Loss values decreased steadily, with the lowest loss recorded at 0.0142. The best model was selected based on the highest validation accuracy using a model checkpoint strategy, as summarized in **Table 1**.

Table 1. Training and Validation Performance per Epoch

Epoch	Training Loss	Training Accuracy	Validation Accuracy
1	0.9072	0.7644	0.2317
2	0.9486	0.8154	0.1463
3	0.9761	0.8347	0.0666
4	0.9764	0.8536	0.0688
5	0.9889	0.7965	0.0339
6	0.9751	0.8464	0.0737
7	0.9895	0.8649	0.0274
8	0.9817	0.7547	0.0531
9	0.9918	0.8677	0.0267
10	0.9955	**0.8689**	0.0142

On the validation dataset, the model achieved an overall accuracy of 86.89%. A detailed classification report revealed that the lymphocyte class attained the highest F1-score of 0.9515, driven by a near-perfect recall of 0.9968. In contrast, the eosinophil class achieved an F1-score of 0.8086, with a precision of 0.7731 and recall of 0.8475, suggesting some confusion with other classes. The macro-averaged values for precision, recall, and F1-score were 0.8738, 0.8690, and 0.8688 respectively, indicating consistent performance across all four classes.

Table 2. Classification Report of the ResNet18

	Precision	Recall	F1-score	Support
Eosinophil	0.7731	0.8475	0.8086	623
Lymphocyte	0.9102	0.9968	0.9515	620
Monocyte	0.9598	0.8081	0.8774	620
Neutrophil	0.8524	0.8237	0.8378	624
Accuracy			0.8689	2487
Macro avg	0.8738	0.8690	0.8688	2487

	Precision	Recall	F1-score	Support
Weighted avg	0.8737	0.8689	0.8687	2487

The confusion matrix in **Figure 3** further illustrates the distribution of predictions. Most eosinophils were correctly classified, but there were 90 instances where neutrophils were misclassified as eosinophils and 65 monocytes that were also predicted as eosinophils. These misclassifications suggest a degree of morphological overlap that affected the model's discrimination capability.

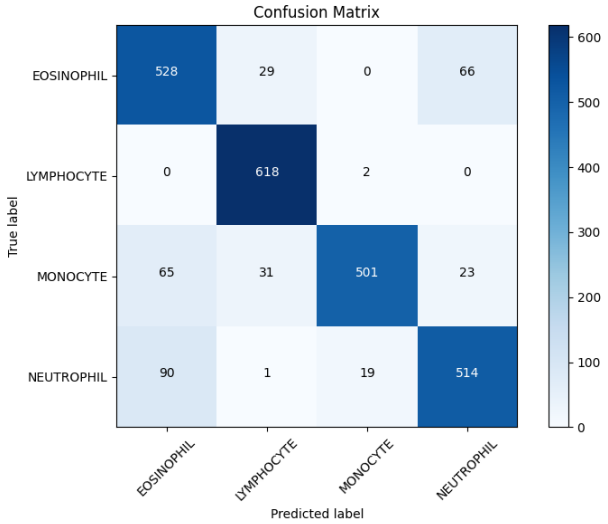


Figure 3. Confusion Matrix of the Best Model.

In addition to categorical metrics, the mean squared error (MSE) and root mean squared error (RMSE) were calculated, yielding 0.7125 and 0.8441, respectively. These values quantify the prediction deviation when treating class labels as ordinal indices.

Discussion

The experimental results demonstrate that ResNet18 is effective for classifying white blood cell images into four distinct types. The model exhibited high recall across all classes, particularly in lymphocytes, indicating its strength in detecting true positives. The comparatively lower precision in the eosinophil class, however, reveals a tendency to misclassify morphologically similar cells such as neutrophils and monocytes. This is further supported by the confusion matrix, where a significant number of neutrophils and monocytes were incorrectly labeled as eosinophils.

Overall, the macro-averaged performance metrics confirm that the model can handle the multi-class classification task with balanced accuracy. The inclusion of MSE and RMSE provides additional insight, capturing classification reliability beyond discrete labels. These error-based metrics suggest that the predicted class indices rarely deviated far from the true classes.

Compared to existing studies that apply deep learning to cell classification, this experiment demonstrates that a baseline architecture such as ResNet18, without additional segmentation modules or ensemble techniques, is sufficient to achieve competitive results. The integration of standard augmentation techniques and transfer learning likely contributed to the model's robustness. These findings affirm the applicability of deep convolutional architectures for haematological image analysis.

4. Conclusion

This study presented a deep learning approach for the multi-class classification of blood cell images using the ResNet18 architecture. The dataset consisted of four leukocyte classes—eosinophils, lymphocytes, monocytes, and

neutrophils—each presenting distinct yet often overlapping morphological features. The proposed model was trained and evaluated using a robust experimental setup, incorporating data augmentation and transfer learning to improve generalization.

The final model achieved a validation accuracy of 86.89%, with a macro-averaged F1-score of 0.8688. Among all classes, lymphocytes were classified with the highest accuracy and recall, while eosinophils posed the greatest challenge due to their morphological similarity to other cell types. The inclusion of multiple evaluation metrics, including MSE and RMSE, allowed for a more nuanced understanding of model performance beyond categorical accuracy.

These results confirm the viability of ResNet18 for automated blood cell classification tasks, highlighting its ability to learn discriminative features from complex medical imagery without requiring elaborate architectural modifications. The findings contribute to the growing body of research on deep learning-based hematological analysis and support further integration of such models into diagnostic workflows.

References:

[1] H. Liu, M. Zhao, C. She, H. Peng, M. Liu, and B. Li, "Classification of CT scan and X-ray dataset based on deep learning and particle swarm optimization," *PLoS One*, vol. 20, no. 1, p. e0317450, Jan. 2025, doi: 10.1371/journal.pone.0317450.

[2] H. Jin, K. Han, H. Xia, B. Xu, and X. Jin, "Detection of weeds in vegetables using image classification neural networks and image processing," *Front. Phys.*, vol. 13, Jan. 2025, doi: 10.3389/fphy.2025.1496778.

[3] C. Ramathal *et al.*, "Deep learning-driven morphology analysis enables label-free classification of therapeutic agentnaive versus resistant cancer cells." Jan. 25, 2025, doi: 10.1101/2025.01.22.634357.

[4] Z. Gao, Y. Tian, S.-C. Lin, and J. Lin, "A CT Image Classification Network Framework for Lung Tumors Based on Pre-trained MobileNetV2 Model and Transfer learning, And Its Application and Market Analysis in the Medical Field," *Appl. Comput. Eng.*, vol. 133, no. 1, pp. 90–96, Jan. 2025, doi: 10.54254/2755-2721/2025.20605.

[5] A. K. Mali, S. Murugappan, J. R. Prasad, S. A. M. Tofail, and N. D. Thorat, "A Deep Learning Pipeline for Morphological and Viability Assessment of 3D Cancer Cell Spheroids." Jan. 24, 2025, doi: 10.1101/2025.01.20.633939.

[6] L. Ma, "Multi-Plant Disease Identification Based on Lightweight ResNet18 Model," *Agronomy*, vol. 13, no. 11, 2023, doi: 10.3390/agronomy13112702.

[7] S. She, "Evaluation of Defects Depth for Metal Sheets Using Four-Coil Excitation Array Eddy Current Sensor and Improved ResNet18 Network," *IEEE Sens. J.*, vol. 24, no. 12, pp. 18955–18967, 2024, doi: 10.1109/JSEN.2024.3367816.

[8] S. T. Prasath, "Deep Learning Framework for Colorectal Cancer Classification Using ResNet18 Based on Dietary Habits Related to Meat Intake and Cooking Methods," *IEEE Access*, vol. 12, pp. 99453–99468, 2024, doi: 10.1109/ACCESS.2024.3430036.

[9] P. Maheswari, "MangoYieldNet: Fruit yield estimation for mango orchards using DeepLabv3 + with ResNet18 architecture," *Multimed. Tools Appl.*, 2025, doi: 10.1007/s11042-025-20791-5.

[10] J. Wang, "Pipeline Landmark Classification of Miniature Pipeline Robot π-II Based on Residual Network ResNet18," *Machines*, vol. 12, no. 8, 2024, doi: 10.3390/machines12080563.

[11] S. Liu, "Fault Diagnosis Strategy Based on BOA-ResNet18 Method for Motor Bearing Signals with Simulated Hydrogen Refueling Station Operating Noise," *Appl. Sci.*, vol. 14, no. 1, 2024, doi: 10.3390/app14010157.

[12] A. Gupta and S. Gupta, "Enhanced Classification of Imbalanced Medical Datasets using Hybrid Data-Level, Cost-Sensitive and Ensemble Methods," *Int. Res. J. Multidiscip. Technovation*, pp. 58–76, Apr. 2024, doi:

10.54392/irjmt2435.

[13] A. Sinra, B. S. W. Poetro, H. Angriani, H. Zein, and ..., "Optimizing Neurodegenerative Disease Classification with Canny Segmentation and Voting Classifier: An Imbalanced Dataset Study,".

[14] Frieyadie, M. Abdullah, and F. A. Setiawan, "The Effectiveness of Resampling Method for Handling Class Imbalances in Software Defect Prediction," in *2023 International Conference on Information Technology Research and Innovation (ICITRI)*, Aug. 2023, pp. 22–27, doi: 10.1109/ICITRI59340.2023.10249255.

[15] J. Sharma, "Deep Learning Approach for Poultry Disease Classification and Early Detection: ResNet18," *2025 International Conference on Pervasive Computational Technologies, ICPCT 2025*. pp. 139–143, 2025, doi: 10.1109/ICPCT64145.2025.10940397.

[16] W. Yang, "Optimizing Facial Expression Recognition: A One-Class Classification Approach Using ResNet18 and CBAM," *Proceedings - 2024 3rd International Conference on Computer Technologies, ICCTech 2024*. pp. 1–5, 2024, doi: 10.1109/ICCTech61708.2024.00009.

[17] J. Zhang, "Prediction of Intrinsically Disordered Proteins Based on Deep Neural Network-ResNet18," *C. - Comput. Model. Eng. Sci.*, vol. 131, no. 2, pp. 905–917, 2022, doi: 10.32604/cmes.2022.019097.

[18] W. Zhang, "Identifying Multiple Apple Leaf Diseases Based on the Improved CBAM-ResNet18 Model under Weak Supervision," *Smart Agric.*, vol. 5, no. 1, pp. 111–121, 2023, doi: 10.12133/j.smartag.SA202301005.

[19] Y. Jia, "Research on Improving ResNet18 for Classifying Complex Images Based on Attention Mechanism," *Communications in Computer and Information Science*, vol. 2139. pp. 123–139, 2024, doi: 10.1007/978-981-97-3948-6_13.

[20] S. Cheng, "Efficient acne classification based on Resnet18," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 13545. 2025, doi: 10.1117/12.3060304.

[21] I. Tahyudin, "High-Accuracy Stroke Detection System Using a CBAM-ResNet18 Deep Learning Model on Brain CT Images," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 788–799, 2025, doi: 10.47738/jads.v6i1.569.

[22] H. S. Obaid, S. A. Dheyab, and S. S. Sabry, "The Impact of Data Pre-Processing Techniques and Dimensionality Reduction on the Accuracy of Machine Learning," in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, Mar. 2019, pp. 279–283, doi: 10.1109/IEMECONX.2019.8877011.

[23] G. Pallavi, "Brain tumor detection with high accuracy using random forest and comparing with thresholding method," *AIP Conference Proceedings*, vol. 2853, no. 1. 2024, doi: 10.1063/5.0198189.

[24] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. 2009.

[25] R. S. Ram, M. V. Kumar, T. M. Al-Shami, M. Masud, H. Aljuaid, and M. Abouhawwash, "Deep Fake Detection Using Computer Vision-Based Deep Neural Network with Pairwise Learning," *Intell. Autom. Soft Comput.*, vol. 35, no. 2, pp. 2449–2462, 2023, doi: 10.32604/iasc.2023.030486.