

Machine Learning Based Normal White Blood Cell Multi-Classification Optimization

Akkireddi Vara Prasad¹, Kalla Rama Krishna^{2*}

¹Associate Professor, Department of Computer Applications, Visakha Institute of Engineering & Technology(A), Visakhapatnam, India

²PG Research Scholar, Department of Computer Applications, Visakha Institute of Engineering & Technology(A), Visakhapatnam, India

Abstract: Identification of various types of white blood cells (WBCs) in a timely accurate manner is essential for diagnosing infections, allergic diseases, and blood disorders via complete blood count (CBC) workflow. Conventional identification relies on manual microscopic analysis of Leishman- or Giemsa-stained blood smears. While accurate when done by an expert, this cumbersome, subjective process relies heavily on trained hematologists who are often scarce in many high-volume or low-resource labs. This study reports the design, implementation, and comparative testing of a deep-learning-based tool used to automatically identify five types of WBCs in blood smears: Basophil, Eosinophil, Lymphocyte, Monocyte, and Neutrophil. Three convolutional neural networks were trained and evaluated in similar conditions; two transfer-learning models, ResNet50V2 and DenseNet121 pre-trained on ImageNet with multiple phases of training and a Custom CNN. All models were trained using a two-stage approach, starting with head-only training with frozen backbone and finishing with selected fine-tuning of identified backbone top layers and data augmentation, batch normalization, dropout, early stopping, and learning-rate scheduling. Interactive desktop graphical user interface was developed to enable one-image classification, compare models' side-by-side, and visualize dataset statistics and model performance in an exploratory manner. The results generated from the experiments performed on the held-out test set indicate that ResNet50V2 achieved the best performance in overall terms scoring 94.88% accuracy in its test and macros. This is higher than DenseNet121 performance of 88.89% and Custom CNN of 85.34%. Basophil was found to be the most unambiguous one to classify, while Eosinophil and Monocyte were ranked as the most difficult classes indicating that particular grades of classification difficulty are bound with meanings that mark certain groups of granulocytes.

Keywords: White Blood Cell Classification, Deep Learning, Convolutional Neural Networks, ResNet50V2, DenseNet121, Transfer Learning, Custom CNN, Medical Image Classification, Complete Blood Count.

1. Introduction

Leukocytes or white blood cells are the special cells of the human immune system. The relative count of different types of leukocytes in blood enables doctors to rapidly assess the immunological and hematological state of a patient [1]. The five major subtypes originating in conventional white blood cell differential counting Lymphocytes, Neutrophils, Basophils, Eosinophils, and Monocytes perform different physiological functions. A change in the proportions of these types are

associated with a number of health problems including different infections as well as allergic and autoimmune diseases [2]. The process of differential cell counting is considered to be “gold standard” when performed through staining of peripheral blood and microscopic examination by a trained technician or doctor. Although effective, this process is time-consuming and labor-intensive and is connected with considerable variability [12].

In the last decade, deep learning, and especially CNN technology have emerged as the most influential technologies in the area of diagnosis in pathology and hematology [20]. Unlike traditional image analysis systems that require the use of manually created visual features like cellular shape, color intensity, CNN needs only access to the pixel data [13]. Transfer learning is a technique that uses a pretrained model from a vast dataset like ImageNet, and then fine-tunes it further to operate in a specific domain of interest. This method is very valuable in many medical imaging applications, where data is rare and costly [19].

The automation of white blood cell classification will prove to be useful in several ways. The use of robots in such classification will shorten the time spent on analyzing a complete blood count, lessen the burden of work on specialists in high-throughput laboratories, and provide hospitals and clinics without proper hematology staff with access to modern diagnostic instruments [16]. It can do much more than just improve accuracy of the predictions: if a system not only offers accurate predictions but also presents important information in an understandable way (i.e. model confidence, per-class results, etc.), the likelihood of its acceptance and widespread usage in the clinical environment will rise [17]. Nonetheless, the choice of network architecture has significant impact on deployment feasibility. Architecture varies greatly in terms of the number of parameters, stability during training, sensitivity to class imbalance, and how much adjustment the system requires to achieve a desired accuracy level. Therefore, systematic and like-for-like comparison of architectures is useful for practitioners who need to determine which model is suitable for their hardware budget [15], [14]. Data preprocessing and augmentation are also important: certain variations concerning staining protocols, lighting, cell orientation, and image resolution in laboratories may influence the performance of models trained on a single dataset without proper augmentation

*Corresponding author: ramakrishnakalla5@gmail.com

and normalization measures taken during training process [18], [11].

In this work, the authors design, implement, and test a complete deep learning pipeline for classification of WBC into five classes, which consists of: (i) data layer responsible for loading and preprocessing labelled blood smear images; (ii) model layer consisting of two types of transfer learning architecture (ResNet50V2 and DenseNet121) and a from-scratch Custom CNN that incorporates both residual and depth wise-separable convolutional blocks; (iii) evaluation layer that measures accuracy, precision, recall, F1-score, confusion matrices, ROC curves, and precision-recall curves on the hold-out testing set. and (iv) a presentation layer implemented as an interactive Tkinter-based graphical user interface that supports single-image classification, side-by-side model comparison, and exploratory data and performance visualization. The comparative results show that ResNet50V2 achieves the strongest overall performance among the three architectures, and the per-class analysis conducted in this work highlights which leukocyte subtypes remain difficult to separate even under a well-tuned transfer-learning regime, thereby motivating targeted future improvements to data collection and model design.

2. Literature Review

A substantial body of prior work has investigated automated classification of white blood cells using both classical machine learning and deep learning techniques, with reported accuracies spanning a wide range depending on dataset size, number of classes, preprocessing pipeline, and model complexity. This section surveys ten representative studies whose reported classification accuracy is lower than the 94.88% test accuracy achieved by the ResNet50V2 model in the present work, situating the present study's comparative, two-phase transfer-learning methodology against this body of literature.

Vatathanavaro *et al.* [3] compared VGG-16 and ResNet-50 for five-class WBC classification. Their best model, ResNet-50, reached 88.3% accuracy, lower than the 94.88% obtained by ResNet50V2 in this study. The CNN and Con-SVM study [4] reported a plain CNN accuracy of 83.91% and a Con-SVM hybrid accuracy of 85.96%, both well below the present result, underlining the benefit of transfer learning over a shallow CNN trained from scratch. The ten-blood-cell-subtype transfer-learning study [5] benchmarked pre-trained VGG, ResNet, Inception, MobileNet, and DenseNet variants, with individual-model accuracies of 91.4%–94.7%, just below the 94.88% obtained here.

The multi-CNN feature-fusion and fuzzy-logic study [6] found standalone MobileNetV2 reaching 93.65% accuracy and standalone ResNet101 only 65.86%, both short of the present ResNet50V2 result. Madhumala *et al.*, as documented in a systematic WBC review [7], used shape descriptors (roundness, form factor, solidity, compactness) with a Naive Bayes classifier, achieving 83.2% accuracy. Habibzadeh *et al.*, also cited in [7], applied K-means clustering with SVM and neural-network classifiers, reporting 80% accuracy.

The global scoping review of leukocyte classification [8]

summarizes Lippeveld *et al.*, who applied Random Forest, Gradient Boosting, ResNet, and DeepFlow to WBC classification, obtaining balanced accuracies of 77.8% and 70.3%. Rosyadi *et al.*, cited in a pathological blood-cell analysis study [9], used Otsu thresholding with k-means clustering and obtained only 67% accuracy, the lowest among the studies surveyed. Yu *et al.*, cited in the same study [9], combined ResNet, InceptionV3, VGG16, and VGG19 and reported 88.5% accuracy. Laddha, also cited in [9], used a pre-trained CNN purely as a feature extractor with an SVM classifier, reaching 93.94% accuracy the closest of the surveyed results to the 94.88% achieved here, yet still lower.

Taken together, these ten studies span hand-crafted feature pipelines, hybrid CNN-SVM approaches, standalone pre-trained backbones, and fusion architectures, with reported accuracies ranging from 67% to 93.94%. Their consistent shortfall relative to the 94.88% accuracy obtained here supports the central premise of this work: a disciplined two-phase transfer-learning protocol extracts more diagnostically useful representations than hand-crafted features or backbones used solely as static feature extractors.

A. Research Gap

Nonetheless, there is still much room for improvement in this field. In particular, comparative research on CNN architectures has encountered problems stemming from variations in training budgets and data augmentation strategies across the architectures. Most research focuses on accuracy achieved on validation datasets, while class-wise performance is often not reported. This prevents proper interpretation of results obtained in similar classes such as eosinophils (10). Since there are virtually no studies employing accessible interactive solutions for single-image inference, model comparison, and dataset analysis, the present investigation aims to fill the gap by training three different models based on the same protocol and common test dataset while providing comprehensive reports of class-based metrics as well as making the models available via the interactive graphical user interface for inference and comparison.

3. Objective of Work

A. Problem Statement

Manual differential counting of white blood cells remains the clinical standard, yet it is slow, dependent on scarce trained personnel, and subject to inter-examiner variability. Existing automated approaches frequently rely on hand-crafted features, use a single model without architectural comparison, or omit per-class detail, limiting their practical trustworthiness. This study addresses the need to design, train, and rigorously compare multiple deep-learning architectures for five-class WBC classification under a shared training protocol, and to expose the resulting models through an accessible interactive tool.

B. Research Objective

The principal aim of this research is to develop, train, and comparatively evaluate deep-learning models for classifying

white blood cells into five categories — Basophil, Eosinophil, Lymphocyte, Monocyte, and Neutrophil. Specific objectives include: (i) preprocessing and augmenting the WBC dataset; (ii) implementing transfer learning using pretrained ResNet50V2 and DenseNet121 backbones alongside a Custom CNN trained from scratch, under a common two-phase training regime; (iii) evaluating all three models on a held-out test set using accuracy, precision, recall, F1-score, confusion matrices, ROC curves, and precision-recall curves; and (iv) delivering the trained models through an interactive GUI for single-image classification, model comparison, and exploratory visualization.

4. Methodology and Architecture

The methodology adopted follows a systematic transfer-learning approach applied uniformly across the three chosen architectures. The image dataset is structured into class-labelled Train and Test folders pertaining to the five types of WBCs. It is initially accessed through an image data pipeline that resizes all images to 224x224 pixels and batches them for easier training.

Data augmentation is carried out during the training process using an augmentation layer that does random horizontal and vertical flipping, rotation of up to 15 degrees and zooming of up to 15% on a real-time basis as it makes each architecture exposed to a wider range of angles and sizes than just raw data. Each of the architectures uses a particular augmentation layer that is available in the preprocessing stage ensuring that augmentation, normalization, and backbone all form a single computation scheme. The ResNet50V2 model is built with ImageNet weights wherein its backbone is frozen during the building process while outfitting it with several elements such as global average pooling, normalization batch, dense layer containing 256 units, the usage of L2 kernel regularization code along with ReLu activation, utilization of dropout set at .4, and softmax for classification. Input data is converted to the $[-1,1]$ format in keeping with the ResNet requirements. The design of the DenseNet121 closely mirrors the ResNet50V2 and uses the DenseNet architecture features to help efficiently carry out training.

The Custom CNN takes the approach of designing a model from scratch based on a compact residual architecture with depth wise separability. Custom CNN consists of convolution and batch normalization layers organized into two standard structures and transformed into a series of blocks with the residual training stages included features are progressively expanded from 128 to 1024 channels, with 1x1 projection shortcuts added at the 256- and 512-channel stages to allow gradient flow around the depth wise blocks. The network concludes with global average pooling, batch normalization, a 256-unit L2-regularized dense layer, a heavier 0.5 dropout rate reflecting the absence of pretrained regularization, and a five-way softmax output.

Training is done in a two-stage process for transfer-learning networks. First step (head training). At this stage, backbone preprocessing is retained, while the classification head is trained via Adam optimizer, being monitored by early stopping

applying validation loss with a patience of seven epochs, adaptive learning for three epochs and decay rate of 0.2, and custom callback methods for early detection of 98.5% of validation accuracy. Second step (fine-tuning). It applies only to ResNet50V2 and DenseNet121 methods, meaning that the top 30 layers of the backbone become available for re-training by compilation of the new network with significantly accomplished learning rate of 1e-5, while garbage collection callback for memory cleaning makes the training process more successful in case of CPU-restricted computing.

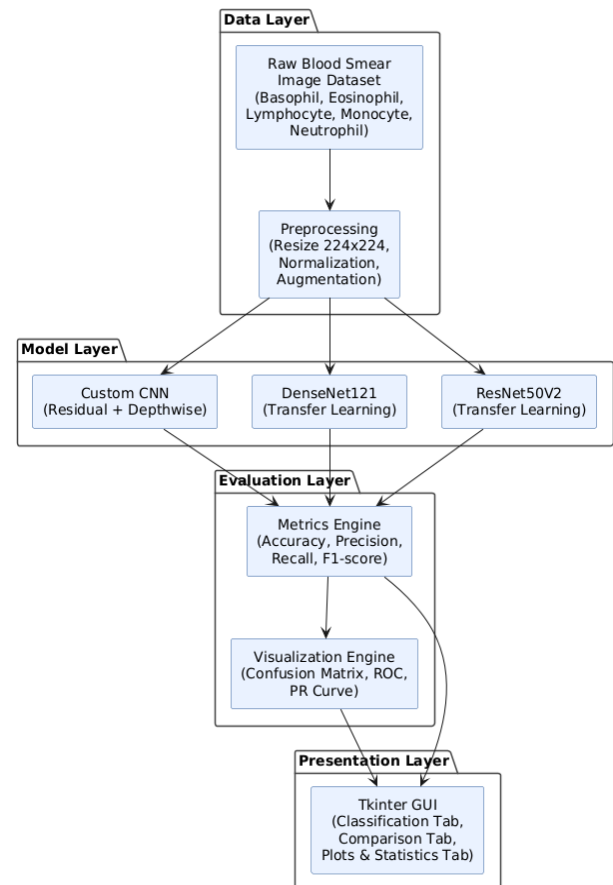


Fig. 1. System architecture diagram

The system is organized in four different modules which follows principles of software engineering. This includes a Data Layer which will be responsible for the loading and resizing datasets. Then comes Model Layer which contains all three convolutional neural network architectures. Resulting accuracy is computed by Evaluation Layer of the system which includes precision, recall, F1 score, confusion matrices, ROC and precision-recall curves. The Presentation Layer is implemented using Tkinter based GUI which provides three tabs, such as Classification which serves single image inference, Model Comparison for comparing and visualizing results across all models, and Plots and Statistics which provides data analysis and visualization of the results obtained.

When a user submits an image through the Classification tab, the GUI first hands the raw image to the preprocessing module, which resizes it to the model's expected input dimensions and

applies the appropriate normalization for the selected architecture. The preprocessed tensor is then passed to the chosen model, which returns a predicted class along with a softmax-derived confidence score. This classification workflow is illustrated in the sequence diagram below.

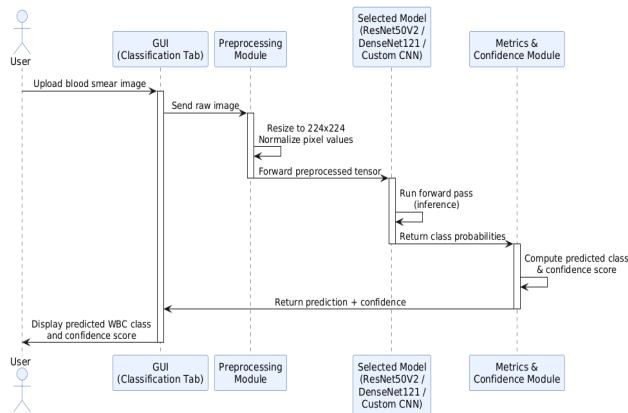


Fig. 2. Classification workflow sequence diagram

The end-to-end training pipeline that produces the deployed models from dataset loading through augmentation, two-phase training, evaluation, and model persistence is summarized in the activity diagram below, which corresponds directly to the implementation of the training script used in this study.

5. Findings and Results

All three architectures were trained and evaluated under the two-phase protocol described above on a held-out test partition that was not used during training or hyperparameter selection. Table 1 summarizes the aggregate test-set performance of the three models.

Table 1

Aggregate test-set performance of the three candidate models

Model	Test Accuracy	Test Loss	Macro F1-score
ResNet50V2	94.88%	0.2904	0.9194
DenseNet121	88.89%	0.4205	0.8330
Custom CNN	85.34%	0.5438	0.7670

ResNet50V2 has obtained maximum overall testing accuracy at a rate of 94.88% with the least test loss of 0.2904 and the maximum macro-averaged F1-score of 0.9194 among all three algorithms meaning it managed to reach significant accuracy while having a good balance between precision and recall among all five WBC classes without having a major preference for some classes while neglecting others. DenseNet121 was able to accomplish a reasonable accuracy result of 88.89% with the macro F1-score of 0.833 and Custom CNN performed the worst, receiving the accuracy of 85.34% and the macro F1-score of 0.767 which means that the effectiveness of transfer learning as a practice has been confirmed as reasonable.

Analysis of the F1-scores for each class showed a similar ordering in all three models. The basophil class was classified the most successfully with F1-scores of 0.9944, 0.9832, and 0.7706 for the ResNet50V2, DenseNet121, and Custom CNN models respectively despite the fact that basophils are usually

the rarest class in the dataset. Lymphocyte has also been classified accurately across models with the F1-scores of 0.9741, 0.8925, and 0.9291 respectively, and the Neutrophils followed with scores of 0.9652, 0.9370, and 0.9023. In contrast, the Eosinophil and Monocyte classes turned out to be the most challenging: the Eosinophil classification gave the results of 0.7908, 0.6839, and 0.5646, and the Monocyte classification yielded the F1-scores of 0.8726, 0.6682, and 0.6682, which is indicative of the morphological similarity of the two classes with their neighbouring granulocyte and agranulocyte classes.

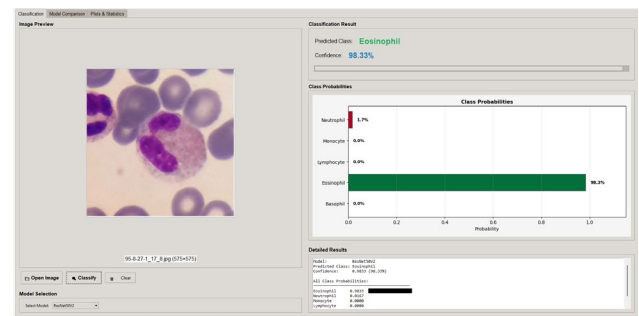


Fig. 3. GUI-Based WBC classification interface: Real-time prediction demo

The Fig. 3 shows the classification application in use, where the ResNet50V2 model is used. The system correctly predicted the class as Eosinophil with 98.33% confidence. The class probability panel confirms strong separation between predictions, with Neutrophil receiving only 1.7% and all other classes near 0%, demonstrating the model's high-confidence, well-calibrated performance on individual samples.

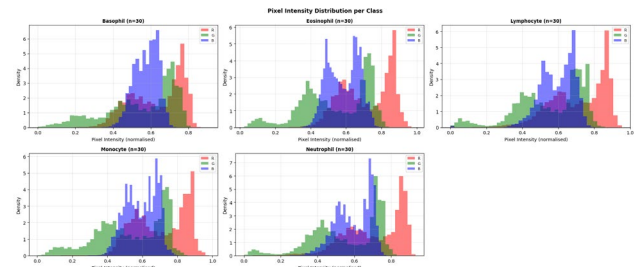


Fig. 4. Pixel intensity distribution across WBC classes

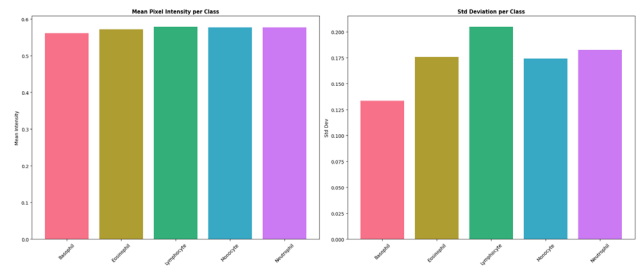


Fig. 5. Mean and Standard Deviation of pixel intensity per class

The Fig. 4 shows RGB pixel intensity distributions ($n=30$ per class) for all five WBC types. Across all classes, the red channel consistently peaks at higher intensity values (0.8–0.9), while blue and green channels dominate lower-to-mid ranges, reflecting typical Giemsa/Wright staining patterns. The distinct

separation between R and B/G channels suggests color intensity is a discriminative feature, potentially contributing to the CNN models' classification performance.

This Fig. 5 compares mean pixel intensity and standard deviation across the five WBC classes. Mean intensities are fairly similar (~ 0.56 – 0.58) across all classes, indicating limited discriminative power from average brightness alone. However, standard deviation varies more noticeably, with Lymphocyte showing the highest variability (~ 0.205) and Basophil the lowest (~ 0.134), suggesting texture/contrast differences may aid classification more than mean intensity alone.

6. Conclusion

The objective of this study was to come up with, develop, and appraise three deep-learning systems: ResNet50V2, DenseNet121, and a Custom CNN on the task of automating the five-class categorization of white blood cells, using blood smear images under an identical two-phase transfer-learning protocol with the same held-out test set. The results showed that ResNet50V2 was the best-performing model with test accuracy of 94.88% and a macro F1-score of 0.9194, achieving much better results than DenseNet121 with 88.89% and a Custom CNN created from scratch with an 85.34% accuracy. It may therefore be concluded from the results that global transfer learning using a large amount of natural image data and two-phase transfer learning approach is better than random initialization approach.

While looking not only at general accuracy, per-class analysis conducted in the current study shows that classifying WBC is not equally difficult for different classes of the cells: for instance, Eosinophil and Monocyte have been reported to be the most difficult to classify whereas Basophil, Lymphocyte, and Neutrophil have been done fairly well by all the three analyzed algorithms. Featuring single-image classification, comparative analysis features for models, and visualization features for features of dataset and working of models, provided interactive GUI shows that precisely hard work with deep learning method can give practical outcome, not only theoretical one.

All findings together indicate that ResNet50V2 used in two-phased transfer learning methodology explained in this article is effective tool for software for automatic WBC differential counting.

7. Limitation

- *Single-source data:* Both training and assessment used only one source of imaging and one system of staining; various procedures within staining, optics, and sample preparation in laboratories may lead to lower performance in real-world data.
- *Class imbalance:* Due to the smaller number of appearances of the basophils and eosinophils compared to neutrophils and lymphocytes, there is a possible explanation for their worse per-class outcome – as they are part of the bigger general accuracy figure.
- *Narrow category coverage:* The research includes only five

large types of white blood cells and does not consider subtypes which can be important in clinical work.

- *Absence of external validation:* All data are based on one isolated test batch without the possibility of any verification of findings against the external data which makes the application in clinical sphere impossible.

8. Future Work

- *Class-balancing techniques:* Apply strategies such as weighted loss functions, focal loss, or targeted oversampling to reduce the performance gap between the strongest and weakest classes.
- *Broader architectural comparison:* Test additional lightweight backbones like MobileNet or EfficientNet to assess whether ResNet50V2's accuracy advantage holds under computational constraints suited for point-of-care or embedded use.
- *Explainability integration:* Incorporate Grad-CAM or similar saliency-based tools to let clinicians visually confirm predictions are based on relevant morphological features rather than image artefacts, boosting clinical trust.
- *External validation:* Conduct prospective validation on independently collected datasets across multiple institutions and staining protocols as a necessary step toward clinical readiness.

References

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [2] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.
- [3] S. Vatahanavaro, S. Tungjitnob, and K. Pasupa, "White blood cell classification using deep learning: VGG-16 and ResNet-50," as cited in "White blood cell classification using custom deep neural network and visualizing features of the images using heatmaps," *Sci. Rep.*, 2026.
- [4] A. Ekiz, K. Kaplan and H. M. Ertunç, "Classification of White Blood Cells Using CNN and Con-SVM," *2021 29th Signal Processing and Communications Applications Conference (SIU)*, Istanbul, Turkey, 2021, pp. 1–4.
- [5] R. Asghar, S. Kumar, and P. Hynds, "Automatic classification of 10 blood cell subtypes using transfer learning via pre-trained convolutional neural networks," *Informatics in Medicine Unlocked*, vol. 49, Art. no. 101542, 2024.
- [6] T. Ullah, K. Ullah, P. Khan *et al.*, "Fuzzy logic and deep learning approach for automated white blood cell detection and classification via multi-CNN feature fusion," *Sci. Rep.*, vol. 15, Art. no. 44546, 2025.
- [7] Madhumala *et al.* and Habibzadeh *et al.*, "White blood cell subtype detection and classification," *arXiv preprint arXiv:2108.04614*, 2021.
- [8] N. Praveen, N. S. Pun, S. K. Sonbhadra, S. Agarwal, M. Syafrullah, and K. Adiyarta, "White blood cell subtype detection and classification," *arXiv preprint arXiv:2108.04614*, 2021.
- [9] Rosyadi *et al.*, Yu *et al.*, and S. Laddha, "Pathological analysis of blood cells using deep learning techniques," *arXiv preprint arXiv:2111.03274*, 2021.
- [10] V. Ranga, S. Gupta, P. Agrawal, and J. Meena, "Pathological Analysis of Blood Cells Using Deep Learning Techniques," *Recent Adv. Comput. Sci. Commun.*, Art. no. e140921185564, Sep. 2020.
- [11] M. Lippeveld *et al.*, "Classification of white blood cells (leucocytes) from blood smear imagery using machine and deep learning models: A global scoping review," *PLOS ONE*, 2024.
- [12] C. Jung, M. Abuhamad, D. Mohaisen, K. Han, and D. Nyang, "WBC image classification and generative models based on convolutional neural network," *BMC Med. Imaging*, vol. 22, no. 1, Art. no. 94, May 2022.

- [13] K. Almezghwi and S. Serte, "Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2020, Art. no. 6490479, Jul. 2020, doi: 10.1155/2020/6490479.
- [14] M. Abou Ali, F. Dornaika, and I. Arganda-Carreras, "White blood cell classification: Convolutional neural network (CNN) and Vision Transformer (ViT) under medical microscope," *Algorithms*, vol. 16, no. 11, Art. no. 525, Nov. 2023, doi: 10.3390/a16110525.
- [15] M. Hussein and F. A. E.-S. Z. El-Mougi, "Integrating deep learning and transfer learning: Optimizing white blood cells classification in medical educational institutions," *J. Big Data*, vol. 12, no. 1, Art. no. 189, 2025, doi: 10.1186/s40537-025-01235-1.
- [16] S. H. Karaddi, H. Bitra, S. S. R. Bairaboina, and B. R. Gudibandi, "White blood cell classification using custom deep neural network and visualizing features of the images using heatmaps," *Sci. Rep.*, vol. 16, no. 1, Mar. 2026.
- [17] P. Jeneessha and V. K. Balasubramanian, "Domain knowledge-infused pre-trained deep learning models for efficient white blood cell classification," *Sci. Rep.*, vol. 15, no. 1, Art. no. 15608, May 2025.
- [18] T. Tamang, S. Baral, and M. P. Paing, "Classification of white blood cells: A comprehensive study using transfer learning based on convolutional neural networks," *Diagnostics*, vol. 12, no. 12, Art. no. 2903, Nov. 2022.
- [19] R. Asghar, S. Kumar, P. Hynds, and A. Shaikat, "Classification of white blood cells using machine and deep learning models: A systematic review," *arXiv preprint arXiv:2308.06296*, Aug. 2023.