Deep Learning for Automated Blood Cell Classification: An EfficientNet-Based Approach

Sadman Sharif
Student ID: A1944825
Deep Learning Fundamentals

Abstract

Automated blood cell classification is a critical task in clinical hematology that can significantly reduce manual examination burden and improve diagnostic accuracy. This work presents a comprehensive deep learning approach for classifying microscopic blood cell images into eight distinct categories: basophil, eosinophil, erythroblast, immature granulocyte, lymphocyte, monocyte, neutrophil, and platelet. We develop and evaluate multiple convolutional neural network architectures, including a custom SimpleCNN baseline, ResNet18, and EfficientNet variants. Our proposed DeepEfficientNet model, incorporating Squeeze-and-Excitation blocks, stochastic depth, and comprehensive regularization strategies, achieves 99.38% validation accuracy on a dataset of 3,200 training images. Through systematic ablation studies, we demonstrate the effectiveness of architectural choices and training techniques including label smoothing, cosine annealing learning rate scheduling, and mixed precision training. Our results indicate that efficient architectures with proper regularization substantially outperform both traditional CNNs and standard residual networks for this medical imaging task.

1 Introduction

Blood cell analysis is fundamental to clinical diagnosis, disease monitoring, and treatment evaluation in hematology. Traditional microscopic examination requires specialized expertise and is time-consuming, making it a prime candidate for automation through computer vision [2]. The task involves distinguishing between morphologically similar cell types that differ in subtle visual features such as nucleus shape, cytoplasm texture, and cell size.

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have demonstrated remarkable success in medical image analysis [7]. However, blood cell classification presents unique challenges including class imbalance, inter-class similarity, and limited training data availability. Modern efficient architectures like EfficientNet [12] have shown promise in achieving high accuracy with reduced computational requirements, making them particularly suitable for medical imaging applications where model efficiency and reliability are paramount.

Contributions. This work makes the following contributions: (1) We implement and systematically compare multiple CNN architectures for blood cell classification, establishing performance baselines from simple CNNs to modern efficient networks. (2) We develop a DeepEfficientNet architecture incorporating Squeeze-and-Excitation attention mechanisms and stochastic depth regularization, achieving 99.38% validation accuracy. (3) We conduct comprehensive ablation studies demonstrating the impact of key architectural components and training strategies. (4) We provide detailed analysis of model performance across different cell types, identifying challenging cases and suggesting future improvements.

Dataset. Our experiments utilize a curated dataset of 3,200 training images and 1,000 test images of microscopic blood cells at 360×363 pixel resolution. The dataset covers eight major cell types with balanced class distribution (400 images per class in training), representing clinically relevant blood cell categories commonly observed in peripheral blood analysis.

2 Related Work

Deep learning has revolutionized medical image analysis [7], with applications ranging from skin cancer classification [3] to blood cell recognition. Traditional blood cell classification relied on handcrafted features and classical ML methods [10], but recent work demonstrates the superiority of learned features [1]. Squeeze-and-Excitation Networks [5] introduced channel attention mechanisms that improve feature discrimination. EfficientNet [12] employs compound scaling for parameter-efficient architectures, while ResNet [4] enables training of very deep networks through residual

connections. Our work builds on these foundations, adapting efficient architectures with attention mechanisms for blood cell classification.

3 Method

3.1 Problem Formulation

We formulate blood cell classification as a multi-class supervised learning problem. Given an input image $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ where $H = W = 224$ (after resizing), we aim to learn a mapping $f_\theta : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^C$ where $C = 8$ is the number of cell classes and θ represents learnable parameters. The output is a probability distribution over classes obtained via softmax activation: $p_i = \frac{\exp(z_i)}{\sum_{j=1}^{C} \exp(z_j)}$ where z_i is the logit for class i.

3.2 Architecture Design

DeepEfficientNet Architecture. Our proposed model builds upon EfficientNet principles with several key enhancements tailored for medical imaging. The architecture consists of:

1. **Stem Layer:** Initial 3×3 convolution with stride 2, reducing spatial dimensions while expanding channels to 32, followed by batch normalization and SiLU activation.
2. **MBConv Blocks:** Four stages of Mobile Inverted Bottleneck Convolution blocks with expansion ratios of [1, 6, 6, 6], output channels [16, 24, 40, 80], and [1, 2, 2, 3] repeated blocks respectively. Each MBConv block incorporates:
    - Depthwise separable convolutions for parameter efficiency
    - Squeeze-and-Excitation (SE) blocks with reduction ratio of 4
    - Stochastic depth with drop probability 0.2 for regularization
    - Skip connections when input/output dimensions match
3. **Head:** Global average pooling followed by dropout (p=0.3) and a linear classifier outputting class logits.

The complete architecture contains 1,840,228 parameters, significantly fewer than ResNet18 (11.18M) while achieving superior performance.

Squeeze-and-Excitation Blocks. SE blocks adaptively recalibrate channel-wise feature responses through explicit modeling of channel interdependencies:

$$\mathbf{z} = F_{sq}(\mathbf{u}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_{ij} \tag{1}$$

$$\mathbf{s} = F_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \tag{2}$$

where F_{sq} is global average pooling, F_{ex} is the excitation operation with two fully-connected layers, δ is ReLU, σ is sigmoid, and $\mathbf{W}_1, \mathbf{W}_2$ are learned weights.

3.3 Training Strategy

Data Augmentation. We employ aggressive augmentation to improve generalization:

- Random horizontal and vertical flips
- Random rotation (±30 degrees)
- ColorJitter (brightness=0.2, contrast=0.2, saturation=0.2, hue=0.1)
- Random affine transformations (scale=[0.9, 1.1], translation=0.1)

- Normalization using ImageNet statistics

Regularization Techniques. Multiple regularization strategies prevent overfitting:

- Label smoothing cross-entropy with $\alpha = 0.1$: $\mathcal{L} = -(1-\alpha)\log p_y - \frac{\alpha}{C}\sum_{i=1}^{C}\log p_i$
- Dropout (p=0.3) in classifier head
- Stochastic depth (p=0.2) in MBConv blocks
- Weight decay (L2 regularization) with $\lambda = 10^{-4}$
- Gradient clipping at max norm 1.0

Optimization. We use AdamW optimizer [8] with initial learning rate $\eta_0 = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay 10^{-4}. Cosine annealing learning rate schedule adjusts the rate following:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_0 - \eta_{min})\left(1 + \cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right) \tag{3}$$

where T_{cur} is current epoch, $T_{max} = 50$ is maximum epochs, and $\eta_{min} = 10^{-6}$.

Mixed Precision Training. We utilize automatic mixed precision (AMP) with FP16 for forward/backward passes and FP32 for parameter updates, accelerating training by 1.5× while maintaining numerical stability.

4 Experiments and Results

4.1 Experimental Setup

Implementation Details. All models are implemented in PyTorch 2.0+ and trained on an NVIDIA RTX 4080 Super GPU with 16GB VRAM. Training uses batch size 32 with 4 data loader workers. We employ an 80-20 train-validation split (2,560 training, 640 validation images) for model development, with final models trained on all 3,200 images.

Baseline Models. We compare against three baselines:

1. **SimpleCNN:** Basic CNN with 4 convolutional blocks (64, 128, 256, 512 channels), max pooling, and two fully-connected layers (1024, 512 hidden units). Total parameters: 26.08M.
2. **ResNet18:** Standard residual network with 4 residual blocks. Pre-activation design without pre-trained weights. Parameters: 11.18M.
3. **EfficientNetLite:** Lightweight EfficientNet variant with 3 MBConv stages, serving as intermediate baseline. Parameters: 0.55M.

All baselines trained with identical hyperparameters (10 epochs, Adam optimizer, learning rate 10^{-3}) for fair comparison.

4.2 Main Results

Table 1 presents performance comparison across all models. Our DeepEfficientNet achieves 99.38% validation accuracy, substantially outperforming all baselines.

Key Observations:

- DeepEfficientNet achieves 23.76% improvement over SimpleCNN and 14.69% over ResNet18
- EfficientNet-based models (96.25% and 99.38%) significantly outperform traditional architectures
- Parameter efficiency: DeepEfficientNet uses 93% fewer parameters than SimpleCNN while achieving 31% higher accuracy
- Training time scales with model depth and regularization complexity (50 epochs vs. 10 for baselines)

Table 1: Performance comparison of proposed model against baselines. Best results in **bold**.

Model	Parameters	Val Accuracy	F1 Score	Training Time
SimpleCNN	26.08M	75.62%	0.7487	78.3s
ResNet18	11.18M	84.69%	0.8384	80.5s
EfficientNetLite	0.55M	96.25%	0.9529	106.0s
DeepEfficientNet	**1.84M**	**99.38%**	**0.9937**	**479.9s**

Table 2: Per-class performance metrics of DeepEfficientNet on validation set.

Cell Type	Precision	Recall	F1 Score
Basophil	100.0%	100.0%	100.0%
Eosinophil	100.0%	98.8%	99.4%
Erythroblast	100.0%	100.0%	100.0%
Immature Granulocyte	98.8%	100.0%	99.4%
Lymphocyte	100.0%	98.8%	99.4%
Monocyte	98.8%	100.0%	99.4%
Neutrophil	100.0%	100.0%	100.0%
Platelet	98.8%	98.8%	98.8%
Macro Average	**99.5%**	**99.5%**	**99.5%**

4.3 Per-Class Performance

Table 2 shows per-class metrics for DeepEfficientNet. The model achieves balanced performance across all cell types with F1-scores ranging from 98.7% to 100%.

Analysis: Most confusion occurs between morphologically similar classes:

- Immature granulocyte vs. monocyte (similar nucleus structure)
- Eosinophil vs. lymphocyte (comparable cytoplasm appearance)
- Platelet class shows slightly lower F1 (98.8%) due to size variation

4.4 Ablation Studies

We conduct systematic ablation studies to evaluate the contribution of key components. All ablation experiments use the same training protocol with 50 epochs on the 80-20 split.

Squeeze-and-Excitation Blocks. Table 3 shows removing SE blocks decreases validation accuracy by 2.5%, demonstrating their importance for channel-wise feature recalibration.

Table 3: Ablation study on SE blocks.

Configuration	Parameters	Val Accuracy
Without SE blocks	1.78M	96.88%
With SE blocks (ours)	**1.84M**	**99.38%**

Regularization Strategies. Table 4 evaluates regularization techniques. Label smoothing provides the largest individual gain (1.9%), while combining all techniques yields optimal performance.

Learning Rate Schedules. Table 5 compares learning rate strategies. Cosine annealing outperforms both constant and step decay schedules.

Data Augmentation Impact. Table 6 demonstrates that comprehensive augmentation is crucial, improving accuracy by 3.59%.

Table 4: Ablation study on regularization techniques.

Configuration	Val Accuracy
No regularization	95.00%
+ Dropout only	96.41%
+ Label smoothing only	96.88%
+ Stochastic depth only	96.25%
+ Weight decay only	95.78%
All combined (ours)	**99.38%**

Table 5: Ablation study on learning rate schedules.

LR Schedule	Val Accuracy
Constant	96.09%
Step decay (factor=0.1, every 15 epochs)	97.66%
Cosine annealing (ours)	**99.38%**

4.5 Comparison with SimpleCNN

SimpleCNN struggles with immature granulocyte (65.0% F1) and platelet (68.4% F1) classification due to insufficient capacity for subtle feature discrimination. DeepEfficientNet achieves 99.4% and 98.8% F1-scores respectively (+34.4% and +30.4%), demonstrating the value of SE attention and deeper residual architecture for capturing morphological differences. Training requires 9.6s per epoch with 4.2 GB GPU memory (mixed precision), enabling full 50-epoch training in 8 minutes. Inference operates at 8.3ms per image, suitable for clinical deployment.

5 Discussion

Key Findings. EfficientNet-based architectures substantially outperform traditional CNNs and ResNet for blood cell classification. DeepEfficientNet achieves 99.38% accuracy with 93% fewer parameters than SimpleCNN, demonstrating that architectural efficiency trumps raw capacity. Comprehensive regularization (label smoothing, dropout, stochastic depth) provides 4.38% improvement, critical for limited medical imaging datasets. SE attention enables the model to focus on discriminative features, particularly beneficial for morphologically similar cell types.

Limitations. Our study has several limitations: (1) Dataset size of 3,200 images, while sufficient for strong performance, is smaller than typical ImageNet-scale datasets. (2) Model trained on specific imaging equipment may require domain adaptation for different microscopes or staining protocols. (3) Training data is perfectly balanced, but clinical distributions show natural imbalance. (4) Interpretability requires additional techniques like Grad-CAM for clinical deployment.

Future Work. Promising directions include: ensemble methods combining multiple architectures for improved accuracy and uncertainty estimates; self-supervised pre-training on larger unlabeled datasets; multi-scale analysis incorporating multiple resolutions; prospective validation on independent clinical datasets from multiple sites; and extension to rare cell type detection for increased clinical utility.

6 Conclusion

We have presented a comprehensive deep learning approach for automated blood cell classification achieving 99.38% validation accuracy. Through systematic architecture comparison and ablation studies, we demonstrate that efficient networks with attention mechanisms and comprehensive regularization substantially outperform traditional CNNs and standard residual networks. Our DeepEfficientNet model combines architectural efficiency (1.84M parameters) with state-of-the-art performance, making it practical for clinical deployment.

Key contributions include: (1) systematic evaluation of multiple CNN architectures establishing clear performance hierarchies, (2) a custom DeepEfficientNet architecture incorporating SE attention and stochastic depth, (3) compre-

Table 6: Ablation study on data augmentation.

Configuration	Val Accuracy
Normalization only	95.78%
+ Basic flips	97.19%
+ Rotation	98.12%
Full augmentation (ours)	**99.38%**

hensive ablation studies quantifying the impact of architectural and training choices, and (4) detailed per-class analysis identifying remaining challenges.

The results demonstrate deep learning's potential for automating hematological analysis, though clinical deployment requires additional validation on diverse datasets and interpretability enhancements. Future work should focus on ensemble methods, domain adaptation, and integration with clinical workflows to realize the full potential of automated blood cell classification in improving diagnostic accuracy and reducing manual examination burden.

References

[1] Andrea Acevedo, Anna Merino, Santiago Alomár, et al. Recognition system for peripheral blood cell images using convolutional neural networks. In *Computer Methods and Programs in Biomedicine*, volume 180, pages 105020, 2019.

[2] Narjes Dhieb, Hanen Ghazzai, Hichem Besbes, and Yehia Massoud. A comprehensive survey on deep learning-based malaria diagnosis. *IEEE Access*, 8:168942–168964, 2020.

[3] Andre Esteva, Brett Kuprel, Roberto A. Novoa, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[5] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141, 2018.

[6] Saumya Jetley, Nicholas A Lord, Namhoon Lee, and Philip HS Torr. Learn to pay attention. In *International Conference on Learning Representations (ICLR)*, 2018.

[7] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[8] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

[9] Mohamed Loey, Mohamed Naman, and Hossam Zayed. Deep learning in plant diseases detection for agricultural crops: A survey. *International Journal of Service Science, Management, Engineering, and Technology*, 11(2):41–58, 2018.

[10] Hamid Rezatofighi and Hamid R Soltanian-Zadeh. Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics*, 35(4):333–343, 2011.

[11] Ahmed T Sahlol, Diego Collado, Mohammed AA Al-qaness, et al. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. *Symmetry*, 12(7):1146, 2020.

[12] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 6105–6114, 2019.