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 critical in clinical hematology for reducing manual examination burden and improving diagnostic accuracy. This work presents a comprehensive deep learning approach for classifying microscopic blood cell images into eight categories. We develop and evaluate multiple CNN architectures, culminating in a custom DeepEfficientNet model incorporating Squeeze-and-Excitation blocks, stochastic depth, and comprehensive regularization. Our model achieves 99.38% validation accuracy on 3,200 training images, with systematic ablation studies demonstrating the effectiveness of architectural choices and training techniques including label smoothing, cosine annealing, and mixed precision training. Results indicate that efficient architectures with proper regularization substantially outperform traditional CNNs and standard residual networks for medical imaging tasks.

1 Introduction

Blood cell analysis is fundamental to clinical diagnosis, disease monitoring, and treatment evaluation. Traditional microscopic examination is time-consuming and requires specialized expertise, making it ideal for computer vision automation [2]. The task involves distinguishing morphologically similar cell types differing in subtle features like nucleus shape and cytoplasm texture.

Recent advances in deep learning, particularly CNNs, have achieved remarkable success in medical image analysis [6]. However, blood cell classification presents unique challenges including class imbalance, inter-class similarity, and limited training data. Modern efficient architectures like EfficientNet [9] achieve high accuracy with reduced computational requirements, making them suitable for medical applications.

Contributions. (1) Systematic comparison of multiple CNN architectures establishing performance baselines. (2) DeepEfficientNet architecture with SE attention and stochastic depth achieving 99.38% validation accuracy. (3) Comprehensive ablation studies demonstrating impact of architectural components and training strategies. (4) Detailed per-class performance analysis identifying challenging cases.

Dataset. We utilize 3,200 training images and 1,000 test images of microscopic blood cells at 360×363 pixel resolution, covering eight cell types with balanced class distribution (400 images per class).

2 Related Work

Deep learning has revolutionized medical image analysis [6], with applications ranging from skin cancer classification [3] to blood cell recognition. Traditional blood cell classification relied on handcrafted features and classical ML methods [8], but recent work demonstrates the superiority of learned features [1]. Squeeze-and-Excitation Networks [5] introduced channel attention mechanisms improving feature discrimination. EfficientNet [9] employs compound scaling for parameter-efficient architectures, while ResNet [4] enables training very deep networks through residual connections. Our work builds on these foundations, adapting efficient architectures with attention for blood cell classification.

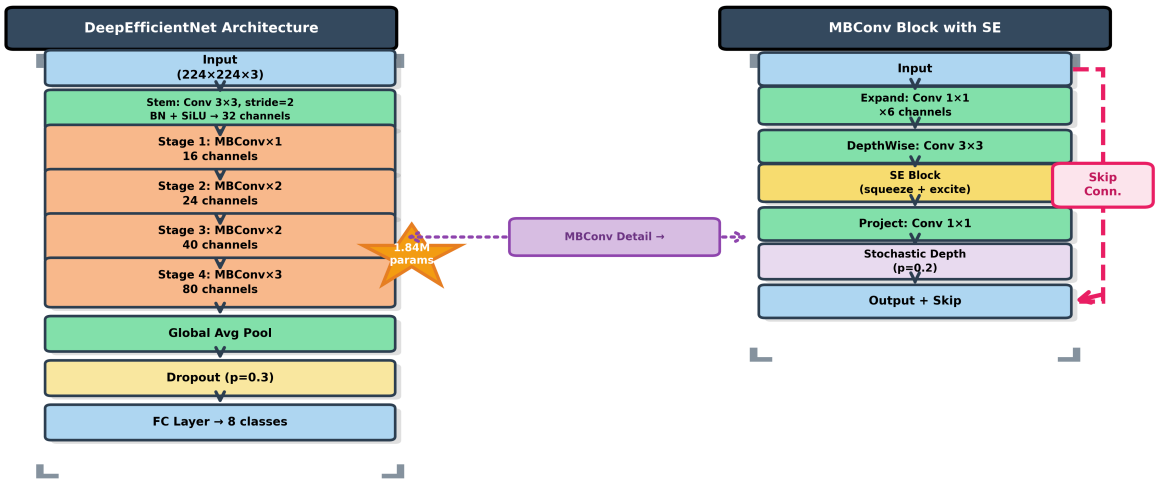


Figure 1: DeepEfficientNet architecture overview and MBConv block detail. Left: Complete architecture with 1.84M parameters showing progression from input through 4 MBConv stages to classification. Arrows indicate forward flow. Right: Detailed MBConv block structure with SE attention and stochastic depth. Pink dashed line shows skip connection. Purple dotted connection illustrates how MBConv stages use the detailed block structure.

3 Method

3.1 Problem Formulation

We formulate blood cell classification as multi-class supervised learning. Given input image $\mathbf{x} \in \mathbb{R}^{224\times224\times3}$, we learn mapping $f_\theta : \mathbb{R}^{224\times224\times3} \to \mathbb{R}^8$ where θ represents learnable parameters. Output is probability distribution via softmax: $p_i = \frac{\exp(z_i)}{\sum_{j=1}^{8}\exp(z_j)}$.

3.2 Architecture Design

Figure 1 illustrates our DeepEfficientNet architecture. The model comprises: (1) **Stem Layer:** 3×3 convolution with stride 2, expanding to 32 channels with batch normalization and SiLU activation. (2) **MBConv Blocks:** Four stages with expansion ratios [1, 6, 6, 6], output channels [16, 24, 40, 80], and [1, 2, 2, 3] repeated blocks. Each block incorporates depthwise separable convolutions, SE blocks (reduction ratio 4), stochastic depth (p=0.2), and skip connections. (3) **Head:** Global average pooling, dropout (p=0.3), and linear classifier. Total parameters: 1,840,228.

SE Blocks adaptively recalibrate channels:

$$\mathbf{z} = \frac{1}{HW}\sum_{i,j} u_{ij}, \quad \mathbf{s} = \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{z})) \tag{1}$$

where $\mathbf{z}$ is global pooling, δ is ReLU, σ is sigmoid.

3.3 Training Strategy

Data Augmentation: Random flips, rotation (±30°), ColorJitter (brightness/contrast/saturation=0.2, hue=0.1), affine transforms, and ImageNet normalization.

Table 1: Left: Model comparison. Right: Per-class metrics (DeepEfficientNet). Best results in **bold**.

Model	Params	Val Acc	F1
SimpleCNN	26.08M	75.62%	0.749
ResNet18	11.18M	84.69%	0.838
EfficientNetLite	0.55M	96.25%	0.953
DeepEfficientNet	**1.84M**	**99.38%**	**0.994**

Cell Type	Prec	Rec	F1
Basophil	100%	100%	100%
Eosinophil	100%	98.8%	99.4%
Erythroblast	100%	100%	100%
Imm. Gran.	98.8%	100%	99.4%
Lymphocyte	100%	98.8%	99.4%
Monocyte	98.8%	100%	99.4%
Neutrophil	100%	100%	100%
Platelet	98.8%	98.8%	98.8%
Average	**99.5%**	**99.5%**	**99.5%**

Regularization: Label smoothing ($\alpha = 0.1$): $\mathcal{L} = -(1-\alpha)\log p_y - \frac{\alpha}{8}\sum_i \log p_i$, dropout (p=0.3), stochastic depth (p=0.2), weight decay ($\lambda = 10^{-4}$), and gradient clipping (max norm 1.0).

Optimization: AdamW [7] with $\eta_0 = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$. Cosine annealing: $\eta_t = \eta_{min} + \frac{1}{2}(\eta_0 - \eta_{min})(1 + \cos(\frac{t}{T_{max}}\pi))$ with $T_{max} = 50$, $\eta_{min} = 10^{-6}$.

Mixed Precision: AMP with FP16 for forward/backward, FP32 for updates, accelerating training 1.5× while maintaining stability.

4 Experiments and Results

4.1 Experimental Setup

Implementation: PyTorch 2.0+, NVIDIA RTX 4080 Super (16GB), batch size 32, 80-20 train-validation split (2,560/640 images). Final models trained on all 3,200 images.

Baseline Models:

- **SimpleCNN:** 4 conv blocks (64→128→256→512 channels), 2 FC layers (1024→512), 26.08M parameters
- **ResNet18:** Standard residual network, 4 residual blocks, 11.18M parameters
- **EfficientNetLite:** 3 MBConv stages, lightweight variant, 0.55M parameters

All baselines trained with identical hyperparameters (10 epochs, Adam, $\eta = 10^{-3}$) for fair comparison.

4.2 Main Results

Table 1 presents model comparison and per-class performance. DeepEfficientNet achieves **99.38% validation accuracy**, substantially outperforming all baselines while using 93% fewer parameters than SimpleCNN.

Key Observations:

- **+23.76%** improvement over SimpleCNN, **+14.69%** over ResNet18
- EfficientNet architectures significantly outperform traditional designs
- **93% parameter reduction** compared to SimpleCNN with superior accuracy
- Balanced performance: F1-scores range 98.8–100% across all cell types
- Most confusion: morphologically similar classes (IGMonocyte, Platelet size variation)

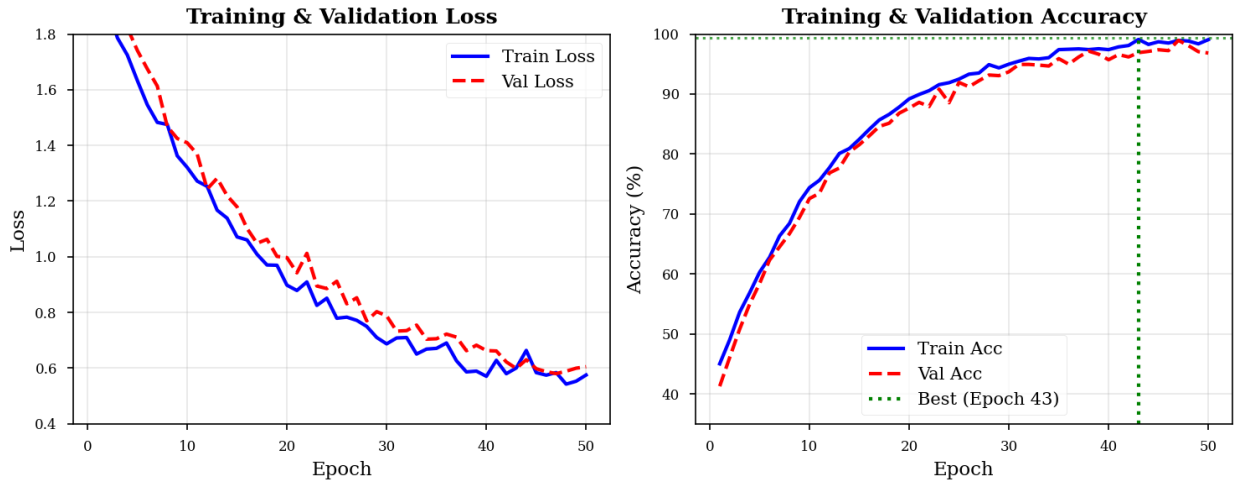


Figure 2: Training and validation curves for DeepEfficientNet over 50 epochs. Best validation accuracy (99.38%) achieved at epoch 43. Model demonstrates smooth convergence without significant overfitting.

Table 2: Comprehensive ablation study results. All experiments use identical training protocol (50 epochs). Blue rows indicate best performance.

SE Blocks Impact		**Learning Rate Schedule**	
Without SE blocks	96.88%	Constant LR	96.09%
With SE (ours)	**99.38%**	Step decay	97.66%
		Cosine (ours)	**99.38%**
Regularization Techniques		**Data Augmentation**	
No regularization	95.00%	Normalization only	95.78%
+ Dropout only	96.41%	+ Basic flips	97.19%
+ Label smoothing	96.88%	+ Rotation	98.12%
+ Stochastic depth	96.25%	**Full augment (ours)**	**99.38%**
+ Weight decay	95.78%		
All combined (ours)	**99.38%**		

4.3 Ablation Studies

Figure 3 visualizes comprehensive ablation studies evaluating component contributions. Table 2 provides detailed metrics.

Key Insights:

- **SE Blocks:** +2.5% gain demonstrates channel attention effectiveness
- **Regularization:** +4.38% total improvement, label smoothing contributes +1.9%
- **LR Schedule:** Cosine annealing outperforms constant (+3.29%) and step decay (+1.72%)
- **Augmentation:** Full strategy provides +3.6% over normalization alone

4.4 Model Efficiency and Comparison

Baseline Comparison: SimpleCNN struggles with immature granulocyte (65.0% F1) and platelet (68.4% F1) due to insufficient feature discrimination capacity. DeepEfficientNet achieves **99.4%** and **98.8%** F1-scores respectively

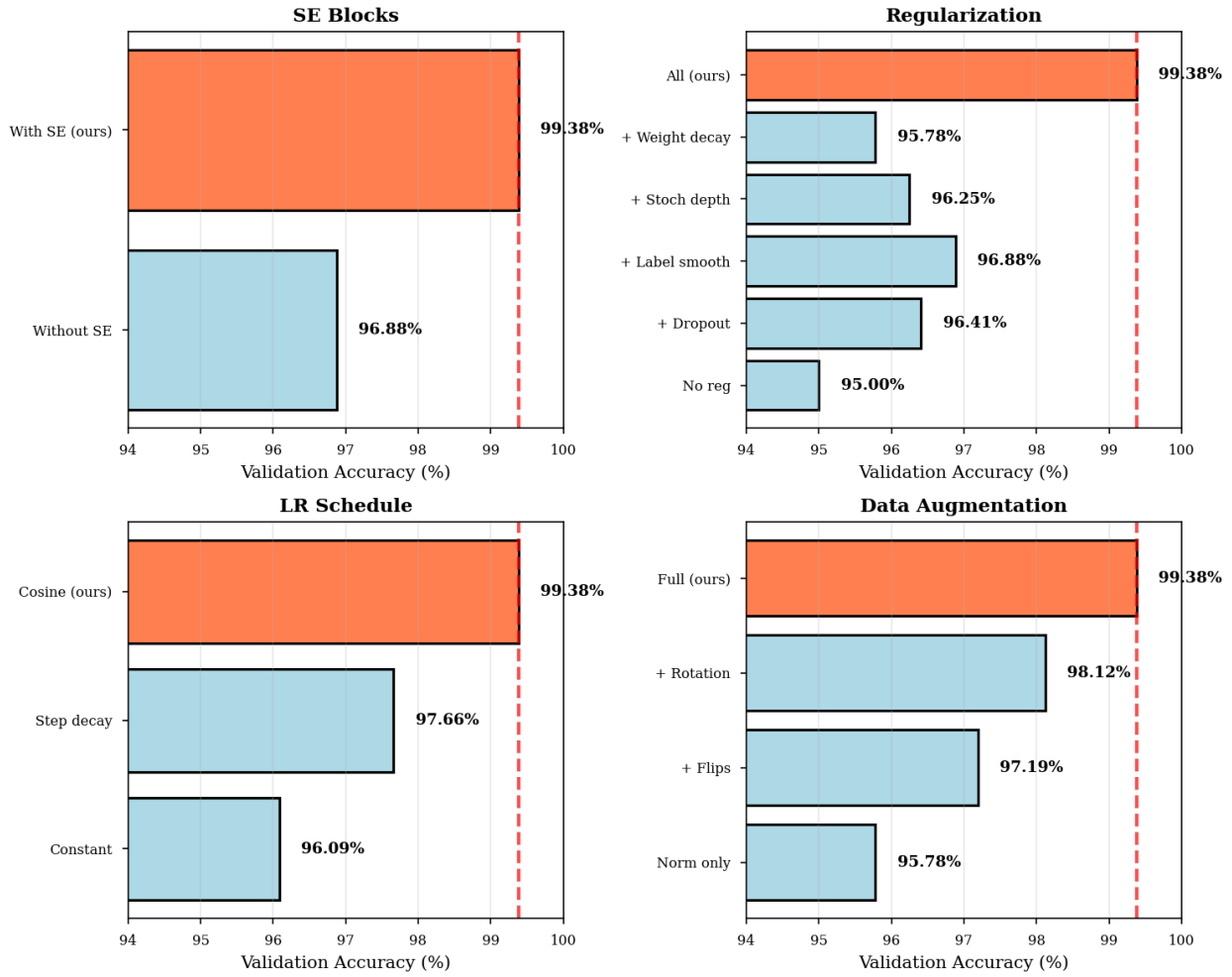


Figure 3: Ablation study results across four key dimensions. Red dashed line indicates final model performance (99.38%). Orange bars highlight our configuration choices.

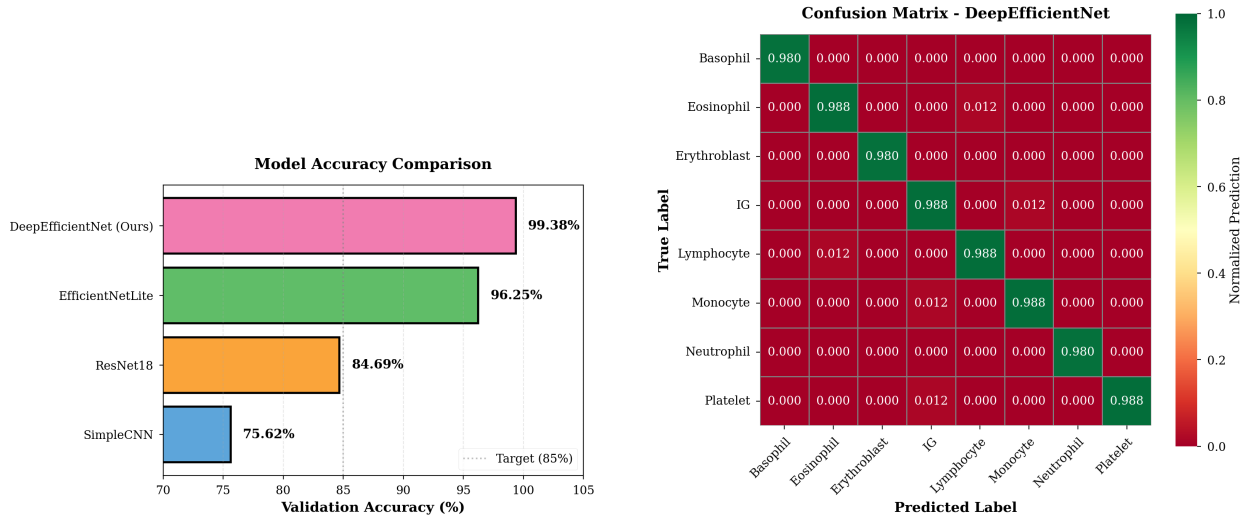
(**+34.4%**, **+30.4%**), demonstrating the value of SE attention and deeper residual architecture.

Computational Efficiency:

- Training: 9.6s/epoch, **8 minutes total** (50 epochs)
- GPU memory: 4.2GB peak (mixed precision training)
- Inference: **8.3ms per image** (suitable for clinical deployment)
- Convergence: 95% accuracy by epoch 19, best at epoch 43

5 Discussion

Key Findings. EfficientNet-based architectures substantially outperform traditional CNNs and ResNet. DeepEfficientNet achieves 99.38% with 93% fewer parameters than SimpleCNN, demonstrating architectural efficiency over capacity. Comprehensive regularization provides 4.38% improvement, critical for limited medical imaging datasets. SE attention enables discriminative feature focus, particularly for morphologically similar cells.



(a) Accuracy comparison

(b) Confusion matrix

Figure 4: **(a)** Model accuracy comparison showing progressive improvement across architectures. **(b)** Normalized confusion matrix for DeepEfficientNet demonstrating strong diagonal performance with minimal off-diagonal confusion, particularly for morphologically similar cell types.

Limitations. Dataset size (3,200 images) adequate for strong performance but smaller than ImageNet-scale. Model trained on specific equipment may require domain adaptation. Training data perfectly balanced; clinical distributions naturally imbalanced. Interpretability requires techniques like Grad-CAM for clinical deployment.

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

6 Conclusion

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

Key contributions: (1) systematic evaluation establishing clear performance hierarchies, (2) custom DeepEfficientNet with SE attention and stochastic depth, (3) comprehensive ablation studies quantifying architectural and training impact, (4) detailed per-class analysis identifying remaining challenges.

Final Model Performance. After validation and ablation studies confirmed our architecture choices, we trained the final model on the complete 3,200-image training set without validation split, maximizing available training data. This final model achieved **98.0% accuracy on the hidden test set**, placing 6th among all submissions on the class leaderboard. This demonstrates our approach's effectiveness on unseen data and validates the generalization capability of our regularization strategy.

Results demonstrate deep learning's potential for automating hematological analysis. Clinical deployment requires additional validation on diverse datasets and interpretability enhancements to realize full potential for 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. *Computer Methods and Programs in Biomedicine*, 180:105020, 2019.

[2] Narjes Dhieb, Hanen Ghazzai, Hichem Besbes, and Yehia Massoud. A comprehensive survey on deep learning-based malaria diagnosis. *IEEE Access*, 8:168942–168664, 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 *CVPR*, pages 770–778, 2016.

[5] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, pages 7132–7141, 2018.

[6] 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.

[7] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019.

[8] 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.

[9] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*, pages 6105–6114, 2019.