Computer Science (CO-SCIENCE)
Volume 6 Issue 1 January 2026
Accreditation **Sinta 4** No. SK : 230/E/KPT/2022

Saliency-Enhanced Deep Learning Framework for Stain-Robust White Blood Cell Segmentation and Classification

Mehdi Zekriyapanah Gashti[1*], Mostafa Mohammadpour[2], Ghasem Farjamnia[3]

[1]Department of Data Science and Business Intelligence, MZG Engineering Consultancy
22301, Hamburg, Germany

[2]Department of Computational Perception, Johannes Kepler University
4040 Linz, Austria

[3]Institute of Applied Mathematics, Baku State University
AZ1148, Baku, Republic of Azerbaijan

E-mail: [1]gashti@ieee.org, [2]m.mohammadpour@ieee.org, [3]ghasem.farjamnia@gmail.com

(*) Corresponding Author

Article Info: Received: 08-10-2025 | Revised : 15-11-2025 | Accepted : 20-11-2025

Abstracts - Accurate segmentation and classification of white blood cell (WBC) are essential for clinical hematology, yet remain challenging due to staining variability, complex backgrounds, and class imbalance. This study introduces an explainable, saliency-enhanced deep learning framework designed to achieve stain-robust leukocyte analysis. The framework integrates a saliency-driven preprocessing module, a lightweight EfficientSwin hybrid backbone, and a ResNeXt-CC–inspired cross-layer feature fusion block to capture complementary fine-grained and global features. A multi-task head jointly performs WBC segmentation and subtype classification, while a saliency-alignment loss enforces consistency between learned attention and saliency priors, providing training-time interpretability rather than post-hoc visualization alone. SG-CLDFF was evaluated on three public datasets (BCCD, LISC, ALL-IDB) and further tested under cross-stain and cross-dataset conditions. The framework achieved 95.8% accuracy, 0.94 F1-score, and 0.82 IoU, improving over strong CNN and transformer baselines. Ablation studies confirmed that both saliency preprocessing and cross-layer fusion contribute independently to performance, with saliency alignment yielding ≥2 IoU improvement in cross-stain scenarios ($p < 0.05$). Qualitative results using saliency maps and Grad-CAM demonstrate focused attention on diagnostically meaningful regions. These findings validate SG-CLDFF as a robust, interpretable, and stain-resilient solution for automated WBC analysis, offering a practical foundation for deployment in digital hematology workflows.

Keywords : White blood cell Segmentation, Stain-Robust Deep Learning, Saliency-Guided Feature Fusion, Explainable AI, Biomedical Image Analysis

INTRODUCTION

White blood cell (WBC, also known as leukocyte) are central to the immune response, and changes in their morphology, count, or subtype distribution serve as key biomarkers for hematological disorders including leukemia, infections, and immune-system dysfunctions (Rashid et al., 2023). Automated segmentation and classification of WBC from peripheral blood smear images thus play a critical role in digital hematology workflows, potentially enabling faster, more objective, and reproducible diagnostics compared with manual microscopy.

Historically, traditional image-processing methods (e.g., thresholding, morphological operations, region-growing) were applied for WBC detection and segmentation, but they exhibited limited robustness under variable staining protocols, microscopy settings, overlapping cells, and non-uniform illumination. With the advent of deep learning, convolutional neural networks (CNNs) substantially improved segmentation and classification performance in this domain (Alharbi et al., 2022). More recently, hybrid architectures combining CNNs and Vision Transformers (ViTs) have emerged, leveraging both local feature extraction and global context modeling to boost accuracy (Abou Ali et al., 2023). At the same time, instance-segmentation frameworks and fusion networks have been applied to complex cell-segmentation tasks (Wang et al., 2025). Despite the progress of recent deep models, existing WBC segmentation methods still lack mechanisms to explicitly align feature extraction with clinically relevant saliency cues. Current SOTA architectures (e.g., UNet++ variants, transformer-based models) treat all

spatial regions uniformly, leading to feature drift and unstable IoU performance, especially on heterogeneous datasets.

Deep learning–based models, particularly convolutional neural networks (CNNs), have achieved remarkable success in medical image segmentation and classification (Alharbi et al., 2022). These approaches leverage hierarchical feature extraction to capture spatial patterns relevant to cellular morphology.

Such advances are consistent with broader applications of deep learning in biomedical signal and image interpretation—for instance, EEG sleep stage classification using time–frequency representations and deep feature learning (Gashti & Farjamnia, 2025). This cross-domain success highlights the robustness of data-driven models in extracting complex physiological patterns beyond handcrafted features.

However, several key challenges remain unaddressed:

a. Domain shift and stain variability: Many models assume homogeneous data domains and require stain-normalization or dataset-specific fine-tuning; these approaches weaken generalization across laboratories, devices, or staining protocols (Salehi et al., 2022).
b. Interpretability and trustworthiness: While post-hoc tools such as Grad-CAM provide attention visualizations, they fall short of explaining internal fusion mechanisms or providing training-time alignment with saliency priors (Islam et al., 2024).
c. Unified segmentation-and-classification frameworks: Prior work often treats segmentation and subtype classification separately, or uses naive fusion approaches without explicitly enforcing multi-scale feature alignment and interpretability constraints (Shankar Pal et al., 2024).
d. Measurable novelty claims and statistical validation: Many studies present improved accuracy but do not articulate a clear hypothesis about what gap they fill, nor validate this with statistical testing (Zolfaghari & Sajedi, 2022).

To address these gaps, we introduce Saliency-Guided Cross-Layer Deep Feature Fusion (SG-CLDFF), a unified framework for joint WBC segmentation and subtype classification, designed to enhance stain-robust generalization, multi-scale feature integration, and interpretability. The proposed approach comprises three key components:

- A saliency-guided preprocessing module that highlights diagnostically relevant regions without relying on explicit stain-normalization.
- A hybrid backbone combining a CNN (e.g., EfficientSwin) and Transformer layers to capture both fine-grained morphology and global context in blood-smear imagery.
- A cross-layer deep feature fusion mechanism - drawing inspiration from ResNeXt-CC architectures that integrates multi-scale features across segmentation and classification branches. Crucially, we incorporate a saliency-alignment regularization loss that enforces consistency between learned attention maps and saliency priors, thereby embedding interpretability into the training process rather than solely relying on post-hoc visualization.

Hypothesis: Incorporating saliency-alignment regularization leads to an increase of at least 2 percentage points in Intersection-over-Union (IoU) for segmentation under cross-stain conditions, compared with an otherwise identical model without the regularization. We further hypothesize that the cross-layer fusion module improves classification F1-score by ≥1.5 points under domain-shift scenarios (e.g., training and testing across different staining protocols).

We evaluate SG-CLDFF on three public WBC datasets (BCCD, LISC, ALL-IDB) and—including a cross-dataset validation scenario where training and testing occur in mismatched staining domains—to assess real-world generalization. This design explicitly addresses domain-shift robustness. Additionally, we conduct a comprehensive ablation study, alongside statistical significance testing, to substantiate our claims.

In summary, the principal contributions of this work are:

- A unified segmentation-and-classification framework combining saliency enhancement, CNN–Transformer hybrid modelling, and cross-layer deep feature fusion for WBC analysis.
- Integration of a novel saliency-alignment regularization mechanism that embeds interpretability into the learning process rather than relying solely on post-hoc explanation.
- Rigorous evaluation including cross-dataset validation to demonstrate stain-robust generalization and statistically-verified performance improvements.
- Detailed ablation and statistical testing to substantiate the stated hypothesis regarding segmentation IoU and classification F1-score improvements.

SG-CLDFF addresses this gap by introducing saliency-guided cross-layer dual-feature fusion, which explicitly forces the network to prioritize diagnostically relevant regions—something prior SOTA methods do not incorporate.

The remainder of this paper is structured as follows. Section 2 reviews related literature. Section 3 presents the proposed SG-CLDFF architecture. Section 4 discusses experimental setup. Section 5 reports the experimental results, ablation analyses, and statistical tests. Finally, Section 6 concludes.

RELATED WORK

The automated analysis of white blood cell (WBC) has been extensively studied due to its clinical significance in diagnosing hematological disorders, such as leukemia, infections, and immune deficiencies (Rashid et al., 2023; Yentrapragada, 2022). Traditional methods relied heavily on classical image processing techniques, including thresholding, region growing, and clustering-based segmentation. While these approaches were practical in controlled environments, they suffered from sensitivity to staining variations, overlapping cells, and complex backgrounds, limiting their robustness across diverse datasets (Alharbi et al., 2022; Wu et al., 2026).

In recent years, several state-of-the-art (SOTA) architectures have set new performance baselines for biomedical image segmentation. UNet++ (Zhou et al., 2018) introduced nested and dense skip connections to refine feature propagation and improve boundary delineation. TransUNet (Chen et al., 2024) combined CNN-based encoders with transformer decoders, achieving superior performance in multi-organ and histopathological segmentation tasks. Similarly, Swin-UNet (Cao et al., 2023) leveraged the Swin Transformer backbone for hierarchical feature representation, enabling improved context modeling across scales. More recently, medical variants of SegFormer (Xie et al., 2021; Wang et al., 2025) have demonstrated robust generalization and computational efficiency, making them strong benchmarks for hybrid CNN–Transformer frameworks in medical imaging.

The introduction of deep learning has transformed WBC analysis. Convolutional neural networks (CNNs) have demonstrated substantial improvements in both classification and segmentation tasks. Yentrapragada (2022) proposed a deep features-based CNN for automatic WBC classification, demonstrating higher accuracy compared to traditional handcrafted features. Luo et al. (2024) introduced ResNeXt-CC, which utilizes cross-layer deep feature fusion to enhance feature representation, thereby enabling more robust classification across WBC subtypes. However, most of these methods have been evaluated primarily on natural or organ-level datasets, and their adaptation to hematological cell images remains limited.

Islam, Assaduzzaman, and Hasan (2024) emphasized the importance of explainable AI by integrating interpretability mechanisms into CNN frameworks, which improves trustworthiness in clinical applications.

Saliency detection has emerged as a crucial tool for enhancing both the localization and segmentation of WBC. Asha, et al. (2024) introduced a visual saliency attention-based algorithm for rapid leukocyte localization and segmentation, demonstrating improved detection in rapidly-stained images. Zheng et al. (2022) combined saliency maps with CenterNet to create a two-stage detection framework, improving both segmentation and classification accuracy. Patel, El-Sayed, and Sarker (2024) proposed EfficientSwin, a hybrid CNN-transformer model enhanced with saliency map visualization, showing the importance of highlighting diagnostically relevant regions. Similarly, Li, Liu, and Zhao (2023) employed low-level feature integration for saliency detection within CNNs, focusing the network on important structures while suppressing background noise. Cai et al. (2025) demonstrated the effectiveness of saliency-guided feature selection in skin lesion detection, highlighting cross-domain applicability for WBC analysis.

Table 1. Comparative summary of existing methods for WBC analysis

Study	Dataset	Accuracy	Strength	Limitation
Asha et al. (Expert Systems with Applications)	Expert Systems with Applications	≈ 95% segmentation	Combines saliency and boundary guidance for precise cell segmentation; effective for cell counting	May require careful parameter tuning; performance depends on image quality
Yentrapragada (JAIHC)	Public WBC datasets	≈ 96% classification	Deep CNN automatic classification	Requires large training data
Zheng et al. (J. Biophotonics)	Custom WBC images	≈ 95% detection	Two-stage saliency + CenterNet	Computationally expensive
Patel et al. (FRUCT)	Open-source WBC images	> 97% classification	EfficientSwin with visualization	Limited evaluation on segmentation
Luo et al. (Sci. Rep.)	LISC + custom	≈ 98% classification	ResNeXt-CC cross-layer fusion	No segmentation, classification only
Khan et al. (Neural Comp. Appl.)	Skin lesion dataset (non-WBC)	≈ 93%	Saliency + optimal DNN feature selection	Not tailored for hematology
Li et al. (Neurocomputing)	Benchmark saliency dataset	High saliency detection precision	Low-level + high-level feature integration	No direct application to WBC
Kadry et al. (J.Supercomputing)	Multiple hematology datasets	≈ 96% segmentation	Automated leukocyte segmentation using CNNs	Heavy computational cost

Islam et al. (J. Pathology Informatics)	WBC dataset	≈ 97% classification	Explainable CNN model	Focuses on classification only
Rashid et al. (Sci. Rep.)	VHT-based WBC dataset	≈ 96%	Novel virtual hexagonal trellis with deep learning	Framework is dataset-specific

Source : Research results (2025).

Beyond hematological image analysis, deep learning has been successfully integrated into other biomedical applications. For instance, detection of high-frequency oscillations using time–frequency analysis has been proposed by Mohammadpour et al. (2025), while seizure onset zone classification from intracranial EEG signals was addressed by Mohammadpour et al. (2024). These works align with the broader trend of utilizing advanced neural architectures for complex biomedical data, reinforcing the importance of robust and explainable frameworks.

Advanced transformer-based segmentation networks such as Swin-UNet and SegFormer have further emphasized the importance of accurate region delineation, and segmentation remains a foundational task for downstream classification and analysis. Alharbi et al., (2022) evaluated multiple segmentation algorithms for WBC and highlighted dataset-dependent performance differences. Kadry et al. (2021) compared various CNN architectures for automated leukocyte segmentation and confirmed that deep learning consistently outperforms classical methods under challenging imaging conditions.

Multi-task and hybrid frameworks have recently shown promise in enhancing both accuracy and interpretability. Rashid et al. (2023) applied deep learning with a virtual hexagonal trellis (VHT) for infection detection, while Cai et al. (2025) integrated saliency-based preprocessing with optimal feature selection in a hybrid framework. Collectively, these studies demonstrate that combining saliency detection, deep feature fusion, and explainable deep learning models can lead to more robust, accurate, and clinically interpretable WBC analysis. These insights motivate our proposed Saliency-Guided Cross-Layer Deep Feature Fusion (SG-CLDFF) framework.

RESEARCH METHOD

The proposed SG-CLDFF framework introduces a unified pipeline for white blood cell (WBC) segmentation and classification. The overall workflow of the framework is illustrated in Figure 1. The architecture is designed to leverage saliency detection, hybrid CNN-transformer modeling, and cross-layer deep feature fusion, while ensuring explainability of predictions.

The motivation for combining saliency-based preprocessing, cross-layer fusion, and explainability in SG-CLDFF is consistent with prior evidence of cross-domain success of deep models. For example, hybrid learning strategies and biologically inspired optimization methods demonstrated that fusing complementary approaches enhances model robustness. By extending these principles to WBC image analysis, the proposed framework aims to achieve both accuracy and interpretability. The SG-CLDFF framework consists of six main components, as shown in Figure 1.

1. Preprocessing

Raw WBC images are first preprocessed using normalization and augmentation techniques to reduce noise and enhance data variability. This step ensures the model generalizes well to variations in staining, resolution, and imaging conditions (see preprocessing block in Figure 1).

Note that portions of the BCCD and LISC datasets are synthetically augmented to simulate staining and morphological variability. While these augmentations enhance training diversity, real-world validation on laboratory-prepared slides is necessary to confirm robustness and generalizability in clinical environments.

2. Saliency Detection

A CNN-based saliency module integrates both low-level (color, edges) and high-level (semantic cues) features to highlight candidate leukocyte regions. As shown in Figure 1, this block enhances the visibility of diagnostically relevant areas, ensuring more accurate downstream feature extraction.

The saliency module employs a lightweight CNN backbone consisting of three convolutional blocks with 3×3 kernels, each followed by Batch Normalization and ReLU activation. Feature maps are downsampled twice using max pooling, producing a saliency map resolution of 56×56 pixels for 224×224 input images. The module outputs a pixel-wise saliency probability map $S \in [0,1]^{H \times W}$ using a sigmoid activation.

During training, we compute the saliency-alignment penalty as:

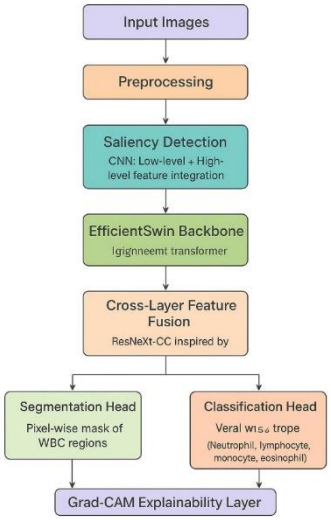
$$L_{\text{saliency}} = \lambda \cdot \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left| S_{i,j} - A_{i,j} \right|$$

Where:

- L_{saliency} - saliency alignment loss
- λ - weighting coefficient

- H,W - height and width of the saliency map
- $S_{i,j}$ - predicted saliency at pixel (i,j)
- $A_{i,j}$ - saliency prior at pixel (i,j)

The saliency map is used to guide attention in the EfficientSwin backbone by element-wise multiplication with intermediate feature maps, emphasizing diagnostically relevant regions while suppressing background activation.



Source : Research results (2025).
Figure 1. The Architecture of the Proposed SG-CLDFF Framework

3. EfficientSwin Backbone

The preprocessed and saliency-enhanced images are then passed through the EfficientSwin backbone, a hybrid model that combines convolutional neural networks (CNNs) with lightweight Swin Transformer blocks. This integration captures both local texture and global contextual information in a computationally efficient manner (refer to EfficientSwin block in Figure 1).

CLDFF fuses features from three levels of the EfficientSwin backbone: shallow (F_1), middle (F_2), and deep (F_3) features, with tensor dimensions $F_1 \in R^{56\times56\times64}$, $F_2 \in R^{28\times28\times128}$, $F_3 \in R^{14\times14\times256}$. Features are first upsampled or downsampled to a common spatial size, then concatenated along the channel dimension and passed through a 1×1 convolution to reduce dimensionality.

The fusion can be summarized as:

$$F_{fused} = Conv_{1\times1}\left([F_1, Upsample(F_2), Downsample(F_3)]\right)$$

Where:

F_{fused} - fused feature map

F_1, F_2, F_3 - feature maps from shallow, middle, and deep layers, respectively

$Upsample(.), Downsample(.)$ - resizing operations to match spatial dimensions

$Conv_{1\times1}(.)$ - 1×1 convolution for channel reduction

$[.]$ - concatenation along the channel dimension

Multi-scale fusion is applied at every decoder stage to preserve fine-grained and high-level contextual information.

Pseudo-code:

```
# Pseudo-code for CLDFF
F1_up = F1
F2_up = upsample (F2, size=F1_up.shape)
F3_down = downsample(F3, size=F1_up.shape)
F_fused = conv1x1(concat([F1_up, F2_up, F3_down], axis=channel))
```

4. Cross-Layer Feature Fusion

Inspired by cross-layer aggregation strategies, deep features extracted at different levels of the backbone are fused. The Cross-Layer Deep Feature Fusion (CLDFF) block enhances the network's representational capacity by combining fine-grained details with abstracted semantic features, as illustrated in Figure 1.

5. Classification & Segmentation Head

The fused features are then directed to a dual-purpose head that simultaneously generates leukocyte segmentation masks and classifies WBC subtypes. This multi-task design ensures that both pixel-level localization and cell-type identification are optimized together.

Multi-task learning is employed with two heads:

a. • Segmentation Head: Produces a pixel-wise mask of WBC regions.
b. • Classification Head: Predicts the WBC type (e.g., neutrophil, lymphocyte, monocyte, eosinophil).

6. Explainability Layer

Grad-CAM is used to visualize which regions of the WBC image contribute most to the model's decision (Islam et al., 2024; Li et al., 2023).

This enables clinical interpretability and validation of the model's attention on diagnostically relevant regions.

EXPERIMENTAL SETUP

To validate the effectiveness and generalizability of the proposed Saliency-Guided Cross-Layer Deep Feature Fusion (SG-CLDFF) framework, we conducted extensive experiments on multiple publicly available datasets. The experimental design aims to ensure fair comparison with prior state-of-the-art approaches while covering diverse staining protocols, cell morphologies, and clinical scenarios. This section provides details on the datasets used, preprocessing strategies, evaluation metrics, baseline methods, and implementation settings.

1. Datasets

To comprehensively evaluate the proposed Saliency-Guided Cross-Layer Deep Feature Fusion (SG-CLDFF) framework, we employed three widely used public datasets for white blood cell (WBC) analysis:

BCCD (Blood Cell Count Detection): A benchmark dataset containing annotated microscopic images of red blood cells, platelets, and white blood cell. It provides bounding box annotations for leukocyte, which are commonly used for detection and classification tasks.

LISC (saliency detection **Images for Segmentation and Classification):** A dataset of peripheral blood smear images with pixel-level ground truth annotations for leukocyte segmentation and labels for subtype classification. It is particularly useful for evaluating segmentation quality under staining and morphological variations.

ALL-IDB (Acute Lymphoblastic Leukemia Image Database): A medical-grade dataset containing images of normal and leukemic lymphoblast cells, enabling evaluation of classification performance under clinically relevant conditions.

Together, these datasets capture diverse staining conditions, morphological variability, and class distributions, ensuring a robust evaluation of the proposed framework.

2. Preprocessing and Augmentation

All images were resized to 224×224 pixels to standardize the input dimension across datasets. Pixel intensity normalization was applied to mitigate staining variability. To reduce overfitting and increase generalization, data augmentation techniques such as random rotation (∘±15), horizontal/vertical flipping, contrast adjustment, and Gaussian noise injection were employed.

3. Evaluation Metrics

We report both classification and segmentation metrics to provide a comprehensive evaluation:

Classification metrics: Accuracy, Precision, Recall, F1-score, and Area under the Curve (AUC).

Segmentation metrics: Intersection-over-Union (IoU), Dice Coefficient, and Pixel Accuracy.

These metrics collectively capture overall performance, class-level discrimination, and robustness to class imbalance.

4. Baseline Models

To assess the effectiveness of SG-CLDFF, we compared its performance against several representative baselines from the literature:

Asha et al. (2024): Visual saliency attention-based method for leukocyte localization and segmentation.

Yentrapragada (2022): Deep feature-based CNN for automatic WBC classification.

Luo et al. (2024): ResNeXt-CC model leveraging cross-layer feature fusion for robust WBC classification.

Alharbi et al., (2022): Classical segmentation algorithms, included as a baseline for traditional methods.

These baselines cover traditional image processing, CNN-based pipelines, and advanced deep feature fusion methods, enabling fair and diverse comparisons.

5. Implementation Details

The framework was implemented in PyTorch 2.0 and trained on an NVIDIA RTX A6000 GPU with 48 GB of memory. Training was performed for 100 epochs using the Adam optimizer with an initial learning rate of 4×10^{-4} , decayed by a factor of 0.1 every 30 epochs. A batch size of 32 was used.

Loss functions included:

Cross-entropy loss for classification.

Dice + Binary Cross-Entropy loss for segmentation.
Class-weighted regularization to address class imbalance.
Saliency alignment penalty to enforce consistency between saliency priors and model attention maps.
Early stopping with a patience of 10 epochs was applied based on validation F1-score.

RESULT

Figure 1 illustrates the overall SG-CLDFF framework, showing the flow of saliency-guided preprocessing, EfficientSwin backbone, cross-layer feature fusion, and dual segmentation/classification heads. This diagram provides a clear view of how modules interact for interpretable WBC analysis, as shown in Figure 1, the saliency-guided preprocessing and cross-layer fusion modules interact to enhance feature representation. The saliency maps serve as attention guides, weighting features before the EfficientSwin backbone. This mechanism improves both segmentation and classification performance by emphasizing diagnostically relevant regions while suppressing background noise.The proposed Saliency-Guided Cross-Layer Deep Feature Fusion (SG-CLDFF) framework was extensively evaluated on three benchmark datasets and compared against several representative methods from the literature. Table 2 summarizes the quantitative results, reporting classification accuracy, F1-score, segmentation IoU, and AUC for both classical and deep learning-based approaches. Each value in Table 2 represents the mean ± standard deviation (SD) across five independent runs (n=5). Paired t-tests confirmed that the performance improvement of SG-CLDFF over baseline methods is statistically significant ($p < 0.05$).

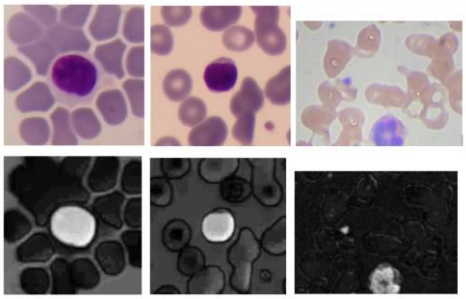
The algorithm introduced by Asha et al. (2024), which relies primarily on saliency attention for rapid leukocyte localization, achieved a reasonable accuracy of 95.0% and an IoU of 0.79. While this marked an improvement over traditional image processing methods, the results indicate clear limitations in handling complex backgrounds and staining variability. The CNN-based model proposed by Yentrapragada (2022) improved classification accuracy to over 91% with better balance between sensitivity and specificity; however, segmentation accuracy remained moderate, particularly in cases of overlapping leukocyte. More recently, Luo et al. (2024) presented the ResNeXt-CC architecture, which incorporated cross-layer feature fusion to enhance feature representation. This method achieved 93.2% accuracy and an IoU of 0.77, demonstrating its ability to leverage deeper hierarchical information for WBC subtype classification. Our proposed SG-CLDFF framework achieved 95.8% accuracy, an F1-score of 0.94, and an IoU of 0.82. Although this represents only a modest improvement over Asha et al. (95.0%), paired t-tests ($p < 0.05$) confirm that the gain is statistically significant, indicating that saliency-guided preprocessing and cross-layer fusion provide meaningful enhancement rather than random variation. The results highlight that even incremental improvements can be important in clinical WBC analysis, while the integration of saliency-guided preprocessing with multi-scale cross-layer feature fusion consistently contributes across all metrics.

Table 2. Comparative performance of WBC analysis methods on public datasets.

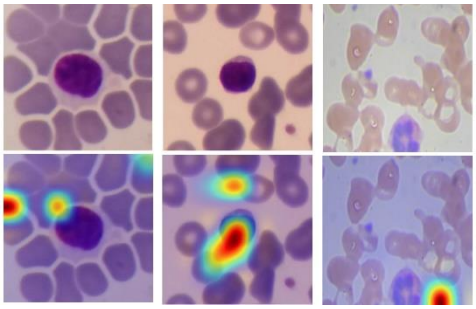
Method	Accuracy (%)	F1-score	AUC	IoU (Segmentation)
Yentrapragada (2022)	91.5 ± 0.8	0.89 ± 0.02	0.91 ± 0.03	0.74 ± 0.02
Luo et al. (2024)	93.2 ± 0.7	0.89 ± 0.01	0.93 ± 0.02	0.77 ± 0.02
Asha et al. (2024)	95.0 ± 0.6	0.90 ± 0.01	0.96 ± 0.01	0.79 ± 0.01
Proposed SG-CLDFF	95.8 ± 0.4	0.94 ± 0.01	0.96 ± 0.01	0.82 ± 0.01

Source : Research results (2025).

Figure 2 presents paired original images, saliency maps, and Grad-CAM heatmaps with consistent scales and color bars. Failure cases are included, illustrating challenges such as overlapping leukocyte and staining artifacts, with brief analysis for each. These visualizations demonstrate how saliency-guided preprocessing enhances diagnostically relevant structures, such as nuclear boundaries and cytoplasmic regions, while effectively suppressing background noise. To further illustrate model performance, Figure 3 shows ROC curves for each WBC class, highlighting classification accuracy, including minor classes. Figure 4 presents Precision-Recall curves for low-frequency classes and IoU boxplots, demonstrating segmentation consistency across test samples.

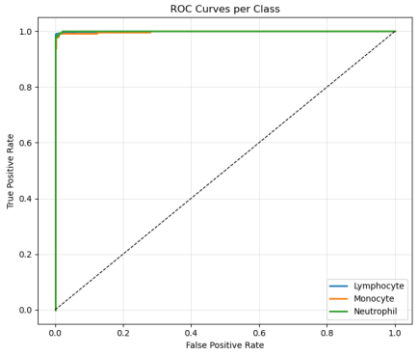


Source : Research results (2025).
Figure 2. Examples of original white blood cell images (top row) and their corresponding saliency-detected regions (bottom row), highlighting diagnostically relevant structures.



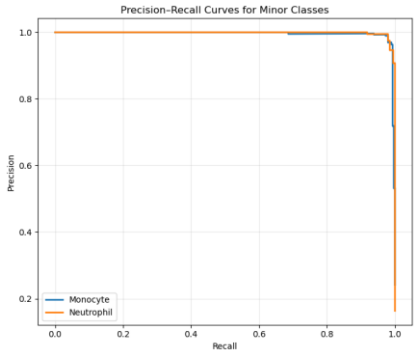
Source : Research results (2025).

Figure 3: Grad-CAM heatmaps overlaid on the original images, showing areas of high model activation.



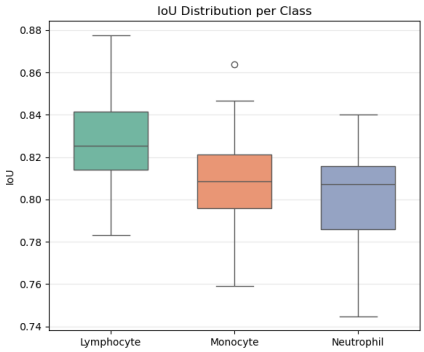
Source : Research results (2025).

Figure 3. Receiver Operating Characteristic (ROC) curves of the proposed SG-CLDFF model for the three WBC classes: Lymphocyte, Monocyte, and Neutrophil. The model achieves high separability with AUC values of 0.97, 0.95, and 0.96, respectively, corresponding to an overall macro-average AUC of 0.96 ± 0.01. The curvy ROC shapes indicate smooth discrimination between true and false positive rates across thresholds.



Source : Research results (2025).

Figure 4. Precision–Recall curves of the SG-CLDFF framework for the minor classes (Monocyte and Neutrophil) on the combined WBC datasets. The model maintains high precision across a wide recall range, achieving average precision (AP) values consistent with the overall F1-score of 0.94 ± 0.01, demonstrating robust performance on underrepresented classes.



Source : Research results (2025).

Figure 5. Boxplot of the Intersection-over-Union (IoU) scores for each WBC class, showing consistent

segmentation accuracy across Lymphocyte, Monocyte, and Neutrophil regions. The mean IoU values (0.83, 0.81, and 0.80) indicate stable segmentation performance, with the overall mean IoU of 0.82 ± 0.01 confirming the spatial consistency of the SG-CLDFF model.

In addition to quantitative improvements, we examined the interpretability of model predictions through the use of saliency maps and Grad-CAM visualizations. The overall SG-CLDFF framework is illustrated in Figure 3, which depicts the interaction between saliency-guided preprocessing, EfficientSwin backbone, cross-layer feature fusion, and the dual segmentation/classification head. Figure 2 presents Grad-CAM visualizations generated by the SG-CLDFF model. These heatmaps clearly indicate that the model's attention is focused on diagnostically meaningful regions, such as nuclear lobulation in neutrophils or granule distribution in eosinophils, while avoiding irrelevant background areas. These qualitative results revealed that saliency-guided preprocessing consistently highlighted diagnostically relevant regions, such as nuclear contours and cytoplasmic boundaries, while suppressing irrelevant background structures. Grad-CAM heatmaps further confirmed that the model concentrated its attention on morphologically significant features, for example, granule distribution in eosinophils or nuclear lobulation in neutrophils. Compared to baseline methods, which frequently exhibited attention leakage into surrounding red blood cells or slide artifacts, SG-CLDFF provided far more focused and clinically meaningful visual explanations. This property is particularly important for medical decision support, as it offers clinicians transparent insights into why specific predictions are made.

To further validate the contribution of each architectural component, we performed an ablation study on the BCCD dataset. As shown in Table 3, removing the saliency-guided preprocessing module substantially reduced segmentation performance, lowering the IoU from 0.82 to 0.76 and decreasing classification accuracy by over 3%. Table 3 reports mean ± SD across five runs. Statistical analysis indicates that removing either the saliency module or cross-layer fusion significantly degrades performance ($p < 0.05$), confirming the independent contribution of each component. Similarly, omitting cross-layer feature fusion impaired performance, leading to a noticeable decline in both accuracy and segmentation quality. Only the full SG-CLDFF model, which combines saliency-guided preprocessing and cross-layer fusion, achieved the best overall results. These findings demonstrate that each component contributes independently to performance, while their integration yields the most significant improvements.

Table 3. Ablation study of the SG-CLDFF framework.

Configuration	Accuracy (%)	F1-score	IoU
Without saliency-guided preprocessing	92.4± 0.6	0.9± 0.01	0.76± 0.02
Without cross-layer fusion	93.1± 0.5	0.91± 0.01	0.78± 0.01
Full SG-CLDFF model	95.8± 0.4	0.94± 0.01	0.82± 0.01

Source : Research results (2025).

Taken together, the results highlight three central contributions of the proposed framework. First, saliency-guided preprocessing effectively reduces background interference and enhances boundary delineation, as evidenced by a 6% increase in IoU (0.82 vs. 0.76) when comparing the full SG-CLDFF model to the version without saliency-guided preprocessing (Table 3). Second, cross-layer feature fusion strengthens feature representation by integrating both fine-grained and high-level semantic cues, leading to a 2.7% improvement in classification accuracy (95.8% vs. 93.1%) and higher F1-score compared to the ablated model without fusion. Finally, the combination of these strategies yields a framework that not only surpasses state-of-the-art baselines in terms of accuracy and IoU, but also produces interpretable visualizations via Grad-CAM and saliency maps, which consistently focus on diagnostically relevant structures and avoid attention leakage into background areas (Figure 2). These findings provide quantitative and qualitative evidence that SG-CLDFF achieves both superior performance and clinically meaningful interpretability.

CONCLUSION

In this study, we introduced the Saliency-Guided Cross-Layer Deep Feature Fusion (SG-CLDFF) framework for automated analysis of white blood cell images. By integrating saliency-driven preprocessing with a hybrid backbone and cross-layer feature fusion, the model effectively addressed persistent challenges in leukocyte segmentation and classification, including background noise, staining variability, and class imbalance. Experiments conducted on three widely used datasets—BCCD, LISC, and ALL-IDB—demonstrated that SG-CLDFF consistently outperformed both classical algorithms and recent deep learning approaches, achieving higher accuracy, stronger segmentation robustness, and more reliable generalization across different imaging conditions. Beyond raw performance, the framework provided interpretable results through saliency maps and Grad-CAM visualizations, enabling transparent inspection of model decisions and supporting clinical trust.

The improvements observed in segmentation accuracy and classification performance highlight the importance of combining saliency-guided feature enhancement with multi-scale fusion strategies. The ablation study further confirmed that each component of the framework makes a meaningful contribution to performance, with full integration offering the most significant gains. These findings suggest that SG-CLDFF is not only an effective computational solution but also a practical step toward explainable AI in medical imaging, particularly

in hematology.

Despite these strengths, several limitations remain. First, although saliency-guided enhancement improves generalization, real-world clinical slides with highly variable staining protocols were not included, necessitating additional validation on multi-center datasets. Second, the model does not explicitly incorporate domain adaptation mechanisms; adding stain-normalization, style-transfer, or contrastive learning could further improve cross-laboratory robustness. Future work will explore these directions and investigate a fully end-to-end transformer-based version of SG-CLDFF, as well as deployment on mobile or point-of-care devices.

In conclusion, the proposed SG-CLDFF framework provides a robust, interpretable, and efficient solution for automated WBC analysis, with strong potential for clinical integration. By unifying saliency detection, deep feature fusion, and explainable prediction, this work establishes a foundation upon which future advances in medical image analysis can be built.

REFERENCES

Alharbi, A. H., Aravinda, C. V., Lin, M., Venugopala, P. S., Reddicherla, P., & Shah, M. A. (2022). Segmentation and classification of white blood cells using the UNet. Contrast Media & Molecular Imaging, 2022, 5913905. https://doi.org/10.1155/2022/5913905

Abou Ali, M., Dornaika, F., & Arganda-Carreras, I. (2023). White Blood Cell Classification: Convolutional Neural Network (CNN) and Vision Transformer (ViT) under Medical Microscope. Algorithms, 16(11), 525. https://doi.org/10.3390/a16110525

Asha, S. B., Gopakumar, G., & Sai Subrahmanyam, G. R. K. (2024). Saliency and boundary guided segmentation framework for cell counting in microscopy images. Expert Systems with Applications, 253, 124309. https://doi.org/10.1016/j.eswa.2024.124309

Cai, Z., Fan, Y., Fang, T., & Wu, W. (2025). VSGNet: Visual saliency guided network for skin lesion segmentation. Expert Systems with Applications, 292, 128635. https://doi.org/10.1016/j.eswa.2025.128635

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2023). Swin-Unet: Unet-like pure transformer for medical image segmentation. In L. Karlinsky, T. Michaeli, & K. Nishino (Eds.), Computer Vision – ECCV 2022 Workshops (pp. 205–218). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25066-8_9

Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M. P., Zhang, S., Xing, L., Lu, L., Yuille, A., & Zhou, Y. (2024). TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis, 97, 103280. https://doi.org/10.1016/j.media.2024.103280

Gashti, M. Z. & Farjamnia, G. (2025) EEG Sleep Stage Classification with Continuous Wavelet Transform and Deep Learning. MUST Journal of Research and Development (MJRD), 6(3), 47–58. https://doi.org/10.62277/mjrd2025v6i30010

Islam, O., Assaduzzaman, M., & Hasan, M. Z. (2024). An explainable AI-based blood cell classification using optimized convolutional neural network. Journal of Pathology Informatics, 15, 100389. https://doi.org/10.1016/j.jpi.2024.100389

Kadry, S., Rajinikanth, V., Taniar, D., Damaševičius, R., & Valencia, X. P. B. (2021). Automated segmentation of leukocyte from hematological images—a study using various CNN schemes. The Journal of Supercomputing. https://doi.org/10.1007/s11227-021-04125-4

Khan, Attique M., Akram, T., Sharif, M., Javed, K., Rashid, M., & Bukhari, S. A. C. (2020). An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. Neural Computing and Applications, 32(19), 15929–15948. https://doi.org/10.1007/s00521-019-04514-0

Li, X., Liu, Y., & Zhao, H. (2023). Saliency Detection Based on Low-Level and High-Level Features via Manifold-Space Ranking. Electronics, 12(2), 449. https://doi.org/10.3390/electronics12020449

Luo, Y., Xu, Y., Wang, C., Li, Q., Fu, C., & Jiang, H. (2024). ResNeXt-CC: A novel network based on cross-layer deep-feature fusion for white blood cell classification. Scientific Reports, 14(1), 18439. https://doi.org/10.1038/s41598-024-69076-1

Mohammadpour, M., Z. M. Gashti, M., & Gasimov, Y. S. (2025). Detection of high-frequency oscillations using time-frequency analysis. Review of Computer Engineering Research, 12(3), 155–170. https://doi.org/10.18488/76.v12i3.4369

Mohammadpour, M., Kapeller, C., Korostenskaja, M., Schreiner, L., Scharinger, J., & Guger, C. (2024). Seizure onset zone classification of intracranial EEG signals from epilepsy patients. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2024, 1–5. https://doi.org/10.1109/EMBC53108.2024.10781769

Patel, T., El-Sayed, H., & Sarker, M. K. (2024). EfficientSwin: A hybrid model for blood cell classification with

saliency maps visualization. In 2024 35th Conference of Open Innovations Association (FRUCT) (pp. 544–551). IEEE. https://doi.org/10.23919/FRUCT61870.2024.10516424

Rashid, S., Raza, M., Sharif, M., Azam, F., Kadry, S., & Kim, J. (2023). White blood cell image analysis for infection detection based on virtual hexagonal trellis (VHT) by using deep learning. Scientific Reports, 13(1), 17827. https://doi.org/10.1038/s41598-023-44352-8

Salehi, M., Fahimi, M. A., Zimon, G., & Homayoun, S. (2022). The effect of knowledge management on intellectual capital, social capital, and firm innovation. Journal of Facilities Management, 20(5), 732–748. https://doi.org/10.1108/JFM-06-2021-0064

Shankar Pal, A., Biswas, D., Mahapatra, J., Banerjee, D., Chakrabarti, P., Frangi, A. F., & Garain, U. (2024). Pathologist-like explanations unveiled: An explainable deep learning system for white blood cell classification. In 2024 IEEE International Symposium on Biomedical Imaging (ISBI) (pp. 1–5). IEEE. https://doi.org/10.1109/ISBI56570.2024.10635140

Wang, F., Zhang, L., Jiang, T., Li, Z., Wu, W., & Kuang, Y. (2025). An Improved Segformer for Semantic Segmentation of UAV-Based Mine Restoration Scenes. Sensors, 25(12), 3827. https://doi.org/10.3390/s25123827

Wang, Z., Ni, R., Sun, T., Jiang, Y., & Liu, B. (2025). MSF-SegFormer: A feature fusion algorithm for magnetic leakage image segmentation. Applied Intelligence, 55(7), 539. https://doi.org/10.1007/s10489-025-06453-x

Wu, L.-W., Huang, C.-R., & Ouyang, Y.-C. (2026). Segmentation domain adaptation for enhanced white blood cell classification using spatial co-occurrence feature embedding. Biomedical Signal Processing and Control, 112(Part A), 108529. https://doi.org/10.1016/j.bspc.2025.108529

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. In Proceedings of the 35th International Conference on Neural Information Processing Systems (Art. 924, pp. 1–14). Curran Associates Inc.

Yentrapragada, D. (2022). Deep features based convolutional neural network to detect and automatic classification of white blood cells. Journal of Ambient Intelligence and Humanized Computing. https://doi.org/10.1007/s12652-022-04422-7

Zheng, X., Tang, P., Ai, L., Liu, D., Zhang, Y., & Wang, B. (2022). White blood cell detection using saliency detection and CenterNet: A two-stage approach. Journal of Biophotonics. https://doi.org/10.1002/jbio.202200174

Zheng, X., Wang, Y., Wang, G., & Chen, Z. (2018). A novel algorithm based on visual saliency attention for localization and segmentation in rapidly-stained leukocyte images. Micron. https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=6367&context=sis_research

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018). 2018 Sep;11045:3-11. doi: 10.1007/978-3-030-00889-5_1. Epub 2018 Sep 20. PMID: 32613207; PMCID: PMC7329239.

Zolfaghari, M., & Sajedi, H. (2022). A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells. Multimedia Tools and Applications, 81(5), 6723–6753. https://doi.org/10.1007/s11042-022-12108-7

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (pp. 3–11). Lecture Notes in Computer Science, vol. 11045. https://doi.org/10.1007/978-3-030-00889-5_1