PRISM: Perinuclear Ring-based Image Segmentation Method for Acute Lymphoblastic Leukemia Classification

Larissa Ferreira Rodrigues Moreira[1], **Leonardo Gabriel Ferreira Rodrigues**[2],
Rodrigo Moreira[1], **André Ricardo Backes**[3]

[1]Institute of Exact and Technological Sciences – Federal University of Viçosa (UFV)
Rio Paranaíba – MG – Brazil

[2]School of Computer Science – Federal University of Uberlândia (UFU)
Uberlândia – MG – Brazil

[3]Departament of Computing – Federal University of São Carlos (UFSCar)
São Carlos – SP – Brazil

`{larissa.f.rodrigues, rodrigo}@ufv.br`

`leonardo.g.rodrigues@ufu.br, andrebackes@ufscar.br`

Abstract. *Automated analysis of peripheral blood smears for Acute Lymphoblastic Leukemia (ALL) is hindered by low contrast and substantial variability in cytoplasmic appearance, which complicate conventional membrane-based segmentation. We found that many recent approaches rely on heavy neural architectures and extensive training, but still struggle to generalize across staining and acquisition variability. To address these limitations, we propose the **P**erinuclear **R**ing-based **I**mage **S**egmentation **Me**thod (PRISM), which replaces explicit cytoplasmic delineation with adaptive concentric zones constructed around the nucleus. These perinuclear regions enable the extraction of robust cytoplasmic descriptors by integrating color information with texture statistics derived from grey-level co-occurrence patterns, without requiring accurate cell-boundary detection. A calibrated stacking ensemble of traditional classifiers leverages these descriptors to achieve a high performance, with an accuracy of 98.46% and a precision-recall AUC of 0.9937.*

1. Introduction

Acute Lymphoblastic Leukemia (ALL) is an aggressive hematological malignancy marked by the rapid proliferation of immature lymphoblasts, requiring timely and accurate diagnosis to guide effective treatment. Although morphological assessment of peripheral blood smears under light microscopy remains the clinical standard, this manual procedure is labor-intensive, subjective, and constrained by the availability of trained hematologists. Computer-Aided Diagnosis (CAD) systems have been proposed to mitigate these limitations, yet extracting key morphological biomarkers such as nuclear enlargement and cytoplasmic reduction remains computationally challenging. Conventional CAD pipelines rely heavily on precise cell membrane segmentation, a step that frequently fails in routine laboratory conditions because of overlapping erythrocytes, low nucleocytoplasmic contrast, and staining inconsistencies [Zolfaghari and Sajedi 2022, Oybek Kizi et al. 2025].

Recent advances in end-to-end Deep Learning (DL), particularly in the field Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), have shifted the focus toward segmentation-free classification [Oybek Kizi et al. 2025]. Despite their strong performance on curated datasets, their clinical adoption is limited by several factors. These models lack transparent decision-making processes, hindering their acceptance in medical environments where interpretability is essential [Aby et al. 2024]. Furthermore, DL systems trained on small medical datasets are prone to overfitting and may inadvertently learn spurious correlations, such as background staining artifacts, rather than true cytological features. Their computational requirements for training and inference also restrict their deployment in resource-limited or decentralized healthcare settings. These challenges highlight the need for diagnostic frameworks that balance accuracy, interpretability, and computational efficiency [Rodrigues Moreira et al. 2025].

To bridge this gap, this paper proposes the **P**erinuclear **R**ing-based **I**mage **S**egmentation **M**ethod (PRISM), a method for hematological feature extraction. Instead of attempting the highly error-prone task of fully segmenting the cell membrane, PRISM leverages the biologically stable basophilic nucleus as a geometric anchor to dynamically generate concentric, adaptive expansion rings. This approach extracts localized morphological, chromatic, and textural gradients from the immediate perinuclear interface and distal cytoplasm, transforming complex cellular pathology into a lightweight, highly discriminative tabular format that systematically rejects the background noise. By coupling this targeted feature extraction with a calibrated ensemble stacking architecture, the proposed approach achieves state-of-the-art diagnostic performance using a fraction of the computational resources. The main contributions of this study are as follows:

- We propose the PRISM, an adaptive mathematical methodology that circumvents the classic bottleneck of full-cell segmentation by analyzing isolated, multiscale perinuclear rings.
- Unlike opaque DL models, PRISM grounds its predictions in explicit, clinically verifiable morphological and chromatic gradients at the nucleocytoplasmic interface.
- Experimental validation demonstrating that the PRISM-driven ensemble successfully outperforms state-of-the-art end-to-end DL architectures on the ALL-IDB2 dataset.
- The translation of a high-dimensional computer vision task into a lightweight tabular machine learning problem, ensuring low-latency inference suitable for CPU-only, resource-constrained clinical environments.

The remainder of this paper is organized as follows. Section 2 provides a review of related studies on automated hematological screening and highlights current methodological gaps. Section 3 details the materials and proposed PRISM method. Section 4 presents and discusses the results. Finally, Section 5 concludes the paper and outlines the future research directions.

2. Related Work

The automated classification of ALL has been dominated by two paradigms: handcrafted feature engineering and end-to-end DL architectures. Early CAD systems prioritized geometric and color properties, such as the nucleocytoplasmic ratio [Scotti 2005]. Subsequent studies explored keypoint descriptors [de Faria et al. 2018], chromatic analysis in the HSV space [Rodrigues et al. 2016], and texture-based descriptors combined with XGBoost [Dias Júnior et al. 2021]. While interpretable, these classic methods often struggle with the morphological variability of leukemic cells [Zolfaghari and Sajedi 2022].

Conversely, DL approaches have achieved high accuracy by leveraging complex architectures [Aby et al. 2024, Oybek Kizi et al. 2025]. Recent works include hybrid ensembles of MobileNetV2, ResNet, and ShuffleNet [Das and Meher 2021, Das et al. 2025], as well as Bayesian CNNs for uncertainty estimation [Hita et al. 2026]. To optimize these models, meta-heuristic strategies such as Elephant Herd [Sahlol et al. 2017] and Salp Swarm [Sahlol et al. 2020] algorithms have been employed. [Rodrigues et al. 2022] attempted to refine this process using a Genetic Algorithm (GA) to optimize a residual network. However, these models function as “black boxes” and impose a significant computational burden, typically requiring high-end GPUs and large datasets for training.

Contribution Positioning. Despite their success, current literature lacks a balance between interpretability and efficiency. The proposed PRISM framework bridges this gap by focusing on the perinuclear interface through a multiscale zonal approach. Unlike DL counterparts, PRISM is a low-cost, GPU-less methodology that provides transparency via verifiable biological markers. By avoiding exhaustive membrane delineation and dense neural computations, it provides an alternative with minimal computational overhead, making it ideal for deployment on resource-constrained diagnostic devices.

3. Material and Methods

The proposed PRISM is a lightweight approach designed for high-throughput ALL screening without GPU dependency. The methodology, illustrated in Figure 1, is organized into three main phases encompassing four integrated steps: (i) preprocessing via nucleus segmentation, (ii) PRISM-based spatial mapping, (iii) multidomain feature extraction, and (iv) meta-classification using calibrated ensemble stacking.

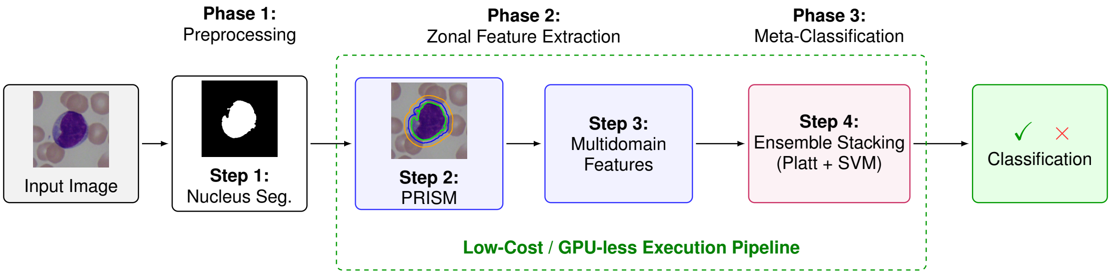

Figure 1. Steps of Proposed Method.

3.1. Dataset

The proposed PRISM method was evaluated using the ALL-IDB2 dataset[1] [Labati et al. 2011], a publicly available benchmark expert-annotated for automated hematological screening. Specifically tailored for single-cell classification tasks, this subset comprises 260 isolated cell images, perfectly balanced between 130 pathogenic leukemic lymphoblasts and 130 healthy lymphocytes (Figure 2).

---

[1] `https://scotti.di.unimi.it/all/`

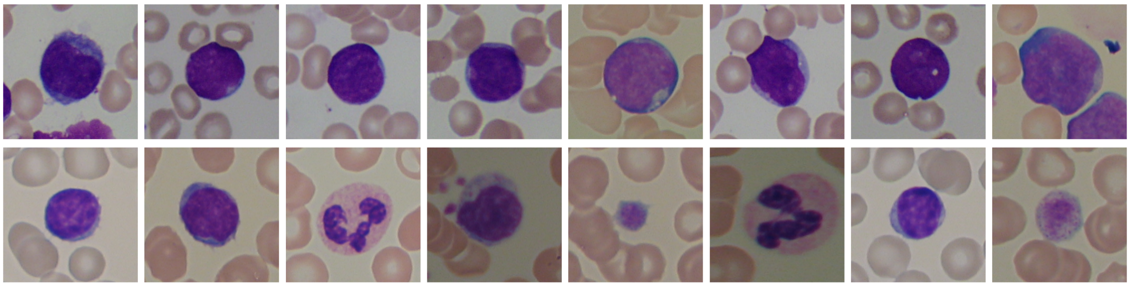

Figure 2. Samples from ALL-IDB2 dataset. Top Row: Pathogenic lymphoblasts. Bottom Row: Healthy lymphocytes.

3.2. Pre-processing and Nuclear Segmentation

To mitigate the illumination and staining inconsistencies inherent to blood smear imaging, the input images were standardized (256×256 pixels) and mapped to the CIELAB color space. The structural definition was enhanced by applying contrast-limited adaptive histogram equalization (CLAHE) exclusively to the lightness (L) channel.

Exploiting the basophilic nature of leukemic nuclei, images were projected into CIELAB and HSV spaces to compute a custom chromatic scoring map, S_{chroma}. This function maximizes the nucleocytoplasmic contrast by inversely weighting the blue-yellow opponent channel (B) and positively integrating saturation (S):

$$S_{chroma} = (255 - B) + 0.5 \times S$$

After Gaussian smoothing, Otsu's thresholding was 75% crop to prevent background bias. For edge cases where extreme staining caused thresholding anomalies (yielding areas $< 1\%$ or $> 65\%$), a robust K-Means clustering fallback ($k = 3$) in the CIELAB space autonomously isolated the darkest, adequately saturated cluster. The binary masks were refined using morphological opening, closing, and hole-filling operations.

Finally, topological heuristic filtering was used to discard artifacts and border-truncated cells. The primary nuclear mask (M_n) was identified by maximizing a multi-objective fitness function ($\mathcal{F}$) that balances solidity, circularity, and mean internal saturation ($\bar{S}$):

$$\mathcal{F} = 0.55 \times \text{Solidity} + 0.35 \times \text{Circularity} + 0.10 \times \left(\frac{\bar{S}}{255}\right)$$

This optimal component, M_n, serves as a geometric anchor for all subsequent spatial extractions.

3.3. PRISM Approach

The main contribution of this study is the PRISM approach (Algorithm 1), which replaces error-prone cell membrane delineation with a multiscale concentric sampling strategy anchored on M_n (Figure 3). To handle morphological variance, PRISM employs an adaptive dilation. Based on the nuclear area A_n and its equivalent radius $R_{eq} = \sqrt{A_n/\pi}$, the expansion radius δ^* is dynamically bounded to prevent overextension: $\delta^* = \min(d, \max(6, \lfloor 1.6 \times R_{eq} \rfloor))$, where $d \in \{10, 24\}$ represents the empirical offsets.

Let $\mathcal{D}_{\delta^*}(M_n)$ denote the morphological dilation with an elliptical structuring element. To isolate the cytoplasmic signatures, two mutually exclusive spatial zones are generated via set difference operations and constrained by a gross cell boundary approximation ($\mathcal{C}$): the proximal interface $Z_1 = (\mathcal{D}_{\delta_1^*}(M_n) \setminus M_n) \cap \mathcal{C}$ and the distal cytoplasmic region $Z_2 = (\mathcal{D}_{\delta_2^*}(M_n) \setminus \mathcal{D}_{\delta_1^*}(M_n)) \cap \mathcal{C}$. These domains ($M_n, Z_1, Z_2$) serve as independent spatial bases for the localized extraction of morphological, chromatic, and textural biomarkers.

Algorithm 1 PRISM: Multiscale Zonal Extraction Pipeline.

```
Require: Enhanced Image $I$, Nuclear Mask $M_n$, Cell Boundary $\mathcal{C}$, Base Radii $d_1, d_2$
Ensure: Multidimensional Feature Vector $V_f$
{Step 1: Adaptive Spatial Decomposition}
$R_{eq} \leftarrow \sqrt{\text{Area}(M_n)/\pi}$
for $i \in \{1, 2\}$ do
    $\delta_i^* \leftarrow \min(d_i, \max(6, \lfloor 1.6 \times R_{eq} \rfloor))$ {Dynamic dilation bounds}
    $E_i \leftarrow \text{Dilate}(M_n, \text{radius} = \delta_i^*)$
end for
$Z_1 \leftarrow (E_1 \setminus M_n) \cap \mathcal{C}$ {Proximal Zone}
$Z_2 \leftarrow (E_2 \setminus E_1) \cap \mathcal{C}$ {Distal Zone}
{Step 2: Feature Engineering & Gradient Computation}
$V_f \leftarrow \emptyset$
for each spatial domain $\mathcal{S} \in \{M_n, Z_1, Z_2, \mathcal{C}\}$ do
    $f_{\mathcal{S}} \leftarrow \text{Extract}(I, \mathcal{S})$ {Morphology, Color Stats, GLCM, LBP}
    $V_f \leftarrow V_f \cup f_{\mathcal{S}}$
end for
$\Delta \leftarrow \text{ComputeGradients}(M_n, Z_1, Z_2)$ {Extract chromatic/textural decay}
$V_f \leftarrow V_f \cup \Delta$
return $V_f$
```

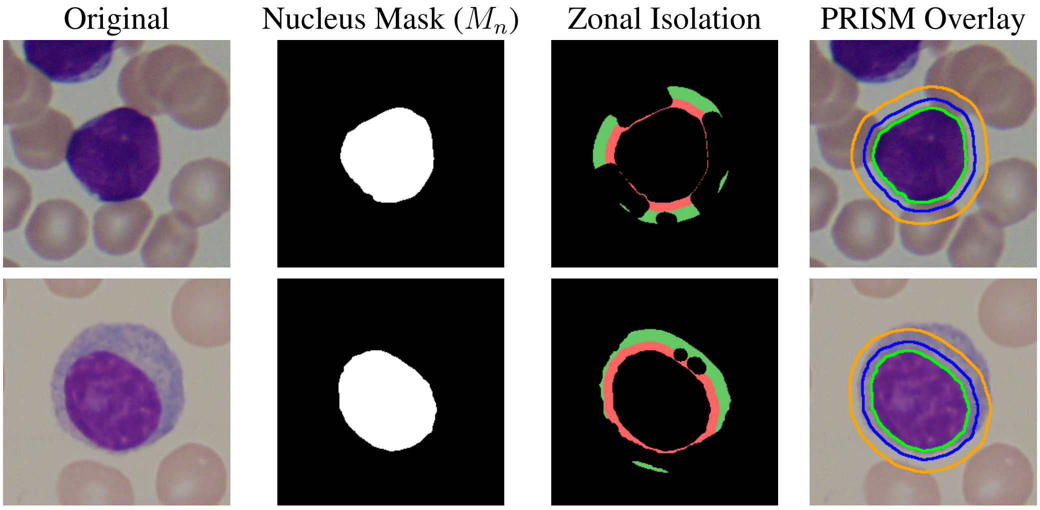

Figure 3. Validation of the PRISM multiscale spatial decomposition across highly pleomorphic hematological profiles. Top row: Pathogenic lymphoblast ALL. Bottom row: Healthy lymphocyte.

3.4. Feature Extraction and Characterization

Quantitative descriptors were extracted from the defined topological domains ($M_n, Z_1, Z_2, \mathcal{C}$) to map leukemogenic alterations into a discriminative feature vector (V_f).

1. Morphological and Topological Descriptors: To quantify nuclear hypertrophy and pleomorphism, geometric properties, including area, circularity, solidity, and boundary roughness, were computed. Topological ratios, specifically the nucleocytoplasmic (N:C) ratio defined as $Area(M_n)/Area(\mathcal{C})$, were calculated to characterize the proportional dominance of the nucleus, a critical clinical marker of ALL.

2. Chromatic Statistics and Spatial Gradients: First-order statistical moments (μ, σ) were extracted for the Red, Blue, Green (RGB) and grayscale channels. A key innovation of PRISM is the computation of spatial intensity gradients (Δ) across the nucleus-ring interfaces:

$$\Delta_{\mu}^{c} = \mu(M_n)_c - \mu(Z_i)_c; \quad \Delta_{\sigma}^{c} = \sigma(M_n)_c - \sigma(Z_i)_c \tag{1}$$

These differential features encode chromatic decay from the nucleus to the cellular periphery, effectively capturing the staining affinity variations characteristic of lymphoblasts.

3. Textural Analysis: To model chromatin distribution, macro-textural properties (Contrast, Homogeneity, Energy, and Correlation) were computed via Gray Level Co-occurrence Matrix (GLCM) for M_n and zones Z_i, averaging results across distances $d \in \{1, 3\}$ and four angular orientations to ensure rotational invariance. In addition, micro-textural heterogeneities were captured using a standard 8-neighborhood Local Binary Pattern (LBP) operator. The resulting 256-bin histograms were downsampled to 32-bin representations via localized summation to mitigate dimensionality and prevent overfitting during the classification stage.

The concatenation of these multidomain attributes yields a robust and interpretable multidimensional vector (V_f) for each cell instance, bridging the gap between raw pixel data and clinical morphology.

3.5. Ensemble Stacking and Probability Calibration

To handle the heterogeneity of V_f, the framework employs a Heterogeneous Ensemble Stacking architecture. This meta-learning strategy combines the diverse inductive biases of multiple base learners to optimize the final decision boundary. The architecture is structured at two levels: Level-0 and Level-1.

Level-0 consists of a diverse pool of base classifiers: Tree-based ensembles Random Forest (RF), Extra Trees (ET), and XGBoost, margin-based models (Support Vector Machine (SVM)), and distance-based algorithms (K-Nearest Neighbors (KNN)). For the distance-sensitive models, the input space was standardized using Z-score normalization. To ensure that the meta-classifier receives reliable inputs, a Probability Calibration stage was integrated. Because algorithms such as the SVM and RF often produce uncalibrated scores, Platt Scaling (sigmoid calibration) was applied via 5-fold cross-validation. This procedure ensures that the Level-0 outputs represent the true posterior probabilities rather than raw decision scores, providing a statistically sound and normalized input space for the meta-learning stage.

Level-1 Meta-Classifier, implemented as a SVM with a Radial Basis Function (RBF) kernel, was trained using an Out-Of-Fold (OOF) strategy to prevent data leakage. During training, 5-fold Stratified Cross-Validation was used to generate OOF probabilities, forming a lower-dimensional meta-feature space. During inference, Level-0 models generate calibrated probabilities, which the Level-1 model synthesizes into the final diagnostic decision, maximizing generalization and mitigating overfitting.

3.6. Evaluation Metrics

All experiments were conducted using a 5-fold Stratified Cross-Validation scheme. For reproducibility, a fixed random seed was applied globally to all stochastic processes, including dataset partitioning and model initialization. In the context of automated oncological screening, standard accuracy is a limited metric. Consequently, the performance was quantified using clinical indicators: sensitivity (True Positive Rate) and specificity (True Negative Rate), which assess the model's reliability in identifying pathological and healthy samples, respectively.

In addition, we used the Matthews Correlation Coefficient (MCC), defined in Equation 2, a statistical rate for binary classification, as it produces a high score only if the model achieves balanced results across all four confusion matrix categories (True Positives, True Negatives, False Positives, and False Negatives).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2}$$

4. Results and Discussion

All experiments were conducted in a virtual machine with 32GB of RAM, $32v$CPU, and an NVIDIA Quadro RTX 6000 GPU on the Fabric testbed [Baldin et al. 2019]. The experiments were programmed using Python 3.12 and PyTorch 3.10. The implementation and all experimental results are publicly available in the GitHub repository[2].

4.1. Performance of the PRISM

To evaluate the discriminative capacity of the multiscale zonal features extracted by the PRISM, an exhaustive search evaluating all 63 possible combinatorial configurations of the base classifiers ($k \in [1, 6]$) was conducted. Base classifiers, comprising RF, ET, SVM, Logistic Regression (LogReg), KNN, and XGBoost, were evaluated as standalone models and in complex multi-model architectures. All stacking combinations utilized a calibrated RBF SVM as the Level-1 meta-classifier.

The experimental results validate the robustness of the morphological, chromatic, and textural signatures extracted. Remarkably, ten distinct stacking configurations achieved the maximum macro-accuracy of 98.46% (MCC $= 0.9698$), demonstrating that the spatial and gradient features inherently mapped the biological nuances of the leukemic domain, independent of a specific algorithm bias.

As detailed in Table 1, while multiple configurations converged to the optimal diagnostic accuracy, the ensemble comprising Extra Trees, SVM, and Logistic Regression (*ET+SVM+LogReg*) emerged as the superior architecture. This specific trio yielded the highest probabilistic stability, achieving a threshold-independent AUC-ROC of 0.9896 and a Precision-Recall AUC (PR-AUC) of 0.9937. The near-perfect PR-AUC is of paramount clinical importance because it indicates an exceptionally low false discovery rate within the minority pathogenic class (lymphoblasts).

---

[2] Available at: `https://github.com/larissafrodrigues/prism-all`

Table 1. Performance Metrics of the Top-Performing Ensemble Stacking Architectures

Level-0 Base Models	Accuracy	Balanced Acc.	MCC	AUC-ROC	PR-AUC
ExtraTrees + SVM + LogReg	**0.9846**	**0.9846**	**0.9698**	**0.9896**	**0.9937**
RF + SVM + LogReg	**0.9846**	**0.9846**	**0.9698**	0.9822	0.9720
RF + LogReg	**0.9846**	**0.9846**	**0.9698**	0.9816	0.9714
ExtraTrees + RF + SVM	**0.9846**	**0.9846**	**0.9698**	0.9813	0.9714
RF + SVM	**0.9846**	**0.9846**	**0.9698**	0.9807	0.9712

4.2. Ablation Study: The Impact of Ensemble Dimensionality

To contextualize the performance gains achieved by the stacking architecture, an ablation study was performed across different ensemble dimensionalities (k). For conciseness, Table 2 summarizes the optimal configurations for each value of k, comparing them against the single base classifier baselines (the complete combinatorial results for all 63 configurations are provided at the `https://github.com/larissafrodrigues/prism-all`).

The data revealed a clear performance plateau, as illustrated in Figure 4. While single classifiers (e.g., Extra Trees) achieved a baseline accuracy of 97.69%, the integration of the meta-classifier was critical for crossing the 98% threshold. Fusing merely two models (*RF + SVM*) immediately maximized the hard-label metrics (98.46% accuracy). Expanding the ensemble to three models optimized the soft-probability metrics (AUC).

However, further increasing the ensemble complexity ($k \geq 4$) yielded no additional accuracy gains, and the complete heavyweight ensemble ($k = 6$) exhibited a slight degradation in performance (98.07% accuracy). This slight performance drop in overly complex ensembles suggests the onset of dimensional overfitting at the meta-classifier level, a phenomenon further corroborated by the variance in the AUC-ROC distribution shown in Figure 5.

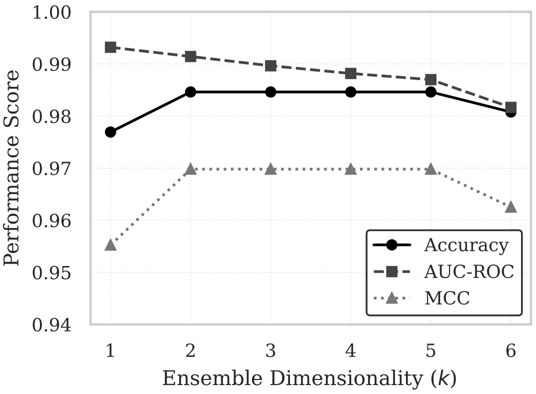

Figure 4. Ablation analysis.

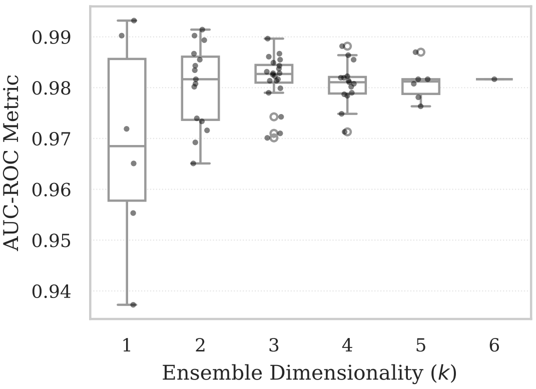

Figure 5. AUC-ROC scores.

4.3. Interpretability and Computational Efficiency

Beyond predictive metrics, the PRISM offers two key diagnostic advantages: using distinct spatial domains (M_n, Z_1, Z_2) ensures classification based on interpretable spatial gradients, contrasting with the opaque "black-box" nature of end-to-end DL paradigms.

Table 2. Comparative Performance of Single Classifiers vs Optimal Stacking Ensembles

Ensemble Size (k)	Model Configuration	Accuracy	Balanced Acc.	MCC	AUC-ROC	PR-AUC
Single Base Classifiers (Level-0 Baselines)						
$k = 1$	Extra Trees (ET)	0.9769	0.9769	0.9552	0.9902	0.9937
$k = 1$	Random Forest (RF)	0.9730	0.9730	0.9478	0.9852	0.9876
$k = 1$	Support Vector Machine (SVM)	0.9653	0.9653	0.9334	0.9893	0.9936
Optimal Ensemble Stacking Configurations (Level-1 Meta: SVM)						
$k = 2$	RF + SVM	**0.9846**	**0.9846**	**0.9698**	0.9807	0.9712
$k = 3$	ET + SVM + LogReg	**0.9846**	**0.9846**	**0.9698**	**0.9896**	**0.9937**
$k = 4$	ET + RF + SVM + LogReg	**0.9846**	**0.9846**	**0.9698**	0.9813	0.9714
$k = 5$	ET + RF + SVM + KNN + LogReg	**0.9846**	**0.9846**	**0.9698**	0.9816	0.9751
Complete Heavyweight Ensemble						
$k = 6$	All Base Classifiers Combined	0.9807	0.9807	0.9625	0.9816	0.9754

In addition, the ablation study proves that a low-latency configuration utilizing merely two base classifiers (*RF + SVM*) is sufficient to preserve the 98.46% diagnostic accuracy while drastically minimizing the computational overhead. This demonstrates that the PRISM framework effectively translates the computationally expensive task of full-cell segmentation into a lightweight, high-performance tabular learning problem, confirming its suitability for rapid clinical screening in resource-constrained environments.

4.4. Comparison with DL Baselines

In contemporary medical image analysis, end-to-end DL models, particularly CNNs and ViTs, are frequently used as the default computational paradigms. To evaluate the PRISM approach, a comparative analysis was conducted with five widely adopted state-of-the-art DL architectures.

The baseline models included classical architectures (ResNet-18 and VGG16), a dense-connectivity model (DenseNet-121), a highly optimized modern CNN (EfficientNet-B0), and a state-of-the-art attention-based model (Swin Transformer - Swin-T). All DL models were trained with Adam optimizer (learning rate =1×10^{-4}, weight decay=1×10^{-4}) and binary cross-entropy with logits for 30 epochs, using a batch size of 16. All models were initialized with ImageNet pre-trained weights and fine-tuned on the raw, unsegmented 256×256 ALL-IDB2 images using the same stratified 5-fold cross-validation scheme to ensure a fair comparison of the results.

As shown in Table 3, the PRISM-driven ensemble outperformed all the DL baselines in terms of accuracy and MCC.

Table 3. Comparison between PRISM and DL Architectures

Methodology	Computational Paradigm	Accuracy	MCC
Swin Transformer (Swin-T)	Vision Transformer (Self-Attention)	0.9192	0.8597
ResNet-18	CNN	0.9654	0.9333
DenseNet-121	CNN	0.9769	0.9557
VGG16	CNN	0.9769	0.9550
EfficientNet-B0	CNN	0.9808	0.9630
PRISM (ET+SVM+LogReg)	**Multiscale Zonal Extraction + Stacking**	**0.9846**	**0.9698**

EfficientNet-B0 achieved the highest performance (98.08% accuracy) among DL models owing to its optimal network balance. Swin-T showed lower performance (91.92% accuracy), typical of self-attention mechanisms that require massive datasets to generalize properly in hematological imaging.

The superiority of the PRISM over these DL paradigms underscores its clinical and computational advantages.

1. **Biological Focus vs. Background Noise:** End-to-end CNNs processing full-frame smears are highly susceptible to learning confounding background artifacts (e.g., erythrocyte density, staining dye variations). PRISM geometrically forces the classifier to evaluate only the relevant biological transition between the nucleus and the proximal/distal cytoplasm.
2. **Interpretability and Hardware Efficiency:** While modern architectures like DenseNet-121 and EfficientNet-B0 act as opaque "black boxes" demanding energy-intensive GPU clusters for training and inference, PRISM effectively reduces the high-dimensional image processing task into a lightweight, tabular Machine Learning problem.

By combining targeted spatial morphology with calibrated ensemble stacking, the PRISM methodology establishes a highly accurate, explainable, and resource-efficient standard for automated lymphoblast screening, outperforming current DL paradigms.

4.5. Comparison with State-of-the-Art

To benchmark the proposed PRISM approach, its classification performance was evaluated against previous studies that used the ALL-IDB2 dataset. The comparative analysis focusing on accuracy (ACC) is summarized in Table 4.

Table 4. Comparison with previous studies.

Author	Methodology	ACC (%)
[Rodrigues et al. 2016]	Hybrid CNN (ResNet + GA)	85.00
[Sahlol et al. 2017]	Handcrafted (EOF+MLP)	91.80
[de Faria et al. 2018]	Handcrafted (SIFT+SURF)	97.22
[Sahlol et al. 2019]	Handcrafted (KNN)	95.67
[Sahlol et al. 2020]	CNN (VGG + SESSA)	96.11
[Das and Meher 2021]	CNN (MobileNet + ResNet)	97.18
[Rodrigues et al. 2022]	CNN (ResNet + GA)	98.46
[Das et al. 2025]	Hybrid CNN Ensemble (MobileNet + ShuffleNet)	98.46
[Hita et al. 2026]	Bayesian CNN + Augmentation	98.65
PRISM (Ours)	**Zonal Stacking Ensemble**	**98.46**

The results demonstrate that PRISM achieves a highly accuracy of 98.46%, matching the performance of complex DL ensembles [Rodrigues et al. 2022, Das et al. 2025] and significantly outperforming traditional handcrafted methods [de Faria et al. 2018, Sahlol et al. 2019]. Although the Bayesian CNN approach by [Hita et al. 2026] reports a higher accuracy, it relies heavily on data augmentation and computationally expensive GPU-accelerated training.

In contrast, PRISM provides a low-cost, GPU-less alternative. By leveraging localized spatial domains (Z_1, Z_2) coupled with an OOF stacking meta-classifier (SVM), our framework extracts highly discriminative signatures directly from the cell morphology. This eliminates the need for deep feature extraction, extensive transfer learning, and exhaustive hyperparameter tuning. Consequently, PRISM reduces training times and operates efficiently in standard CPU environments, establishing itself as a highly scalable solution for resource-constrained clinical screenings.

5. Conclusion

Accurate automated screening of ALL is often hindered by low contrast, cytoplasmic variability, and the brittleness of full-cell membrane segmentation. This study presents PRISM, an interpretable framework that leverages stable nuclear geometry and extracts chromatic and textural cues across adaptive perinuclear zones to capture the morphology linked to leukemic transformation. Combined with a calibrated stacking ensemble, PRISM achieved a strong performance on ALL-IDB2, reaching an accuracy of 98.46% and an MCC of 0.9698.

Comparative analyses indicate that PRISM performs on par with and, in several cases, surpasses contemporary DL models while remaining lightweight and transparent, enabling fast inference without specialized hardware in resource-limited or decentralized settings. Future studies will integrate PRISM into portable or edge-based microscopy pipelines to support accessible point-of-care diagnostics. In addition, we plan to validate additional datasets, such as real-world clinical data from different laboratories. Overall, the results show that segmentation-light, interpretable representations can deliver high-accuracy ALL screening with practical efficiency and clearer decision-making support.

Acknowledgments

The authors gratefully acknowledges the financial support of FAPEMIG (Grant #APQ00923-24). André R. Backes gratefully acknowledges the financial support of CNPq (Grant #302790/2024-1).

References

Aby, A. E., Salaji, S., Anilkumar, K., and Rajan, T. (2024). A review on leukemia detection and classification using artificial intelligence-based techniques. *Computers and Electrical Engineering*, 118:109446.

Baldin, I., Nikolich, A., Griffioen, J., Monga, I. I. S., Wang, K.-C., Lehman, T., and Ruth, P. (2019). FABRIC: A national-scale programmable experimental network infrastructure. *IEEE Internet Computing*, 23(6):38–47.

Das, P. K. and Meher, S. (2021). An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. *Expert Systems with Applications*, 183:115311.

Das, P. K., Meher, S., Rath, A., and Panda, G. (2025). An efficient deep learning system for automatic detection of acute lymphoblastic leukemia. *ISA Transactions*, 158:488–496.

de Faria, L. C., Rodrigues, L. F., and Mari, J. F. (2018). Cell classification using handcrafted features and bag of visual words. In *Anais do XIV Workshop de Visão Computacional*, pages 68–73.

Dias Júnior, D., Cruz, L., Diniz, J., Júnior, G. B., and Silva, A. (2021). Classificação automática de glóbulos brancos usando descritores de forma e textura e extreme gradient boosting. In *Anais do XXI Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 95–106, Porto Alegre, RS, Brasil. SBC.

Hita, X., Javed, F., and Lodi, S. (2026). Reliable leukemia detection via transfer-enhanced bayesian cnns. *Computers in Biology and Medicine*, 202:111419.

Labati, R. D., Piuri, V., and Scotti, F. (2011). All-idb: The acute lymphoblastic leukemia image database for image processing. In *2011 18th IEEE International Conference on Image Processing*, pages 2045–2048.

Oybek Kizi, R. F., Theodore Armand, T. P., and Kim, H.-C. (2025). A review of deep learning techniques for leukemia cancer classification based on blood smear images. *Applied Biosciences*, 4(1).

Rodrigues, L. F., Backes, A. R., Travençolo, B. A. N., and de Oliveira, G. M. B. (2022). Optimizing a deep residual neural network with genetic algorithm for acute lymphoblastic leukemia classification. *Journal of Digital Imaging*, 35(3):623–637.

Rodrigues, L. F., Silva, J. H., Gondim, P. H. C. C., and Mari, J. F. (2016). Leukocytes classification in microscopy images for acute lymphoblastic leukemia identification. In *XII Workshop de Visão Computacional*, pages 68–73.

Rodrigues Moreira, L. F., Moreira, R., Travençolo, B. A. N., and Backes, A. R. (2025). Deep learning based image classification for embedded devices: A systematic review. *Neurocomputing*, 623:129402.

Sahlol, A. T., Abdeldaim, A. M., and Hassanien, A. E. (2019). Automatic acute lymphoblastic leukemia classification model using social spider optimization algorithm. *Soft Computing*, 23(15):6345–6360.

Sahlol, A. T., Ismail, F. H., Abdeldaim, A., and Hassanien, A. E. (2017). Elephant herd optimization with neural networks: A case study on acute lymphoblastic leukemia diagnosis. In *2017 12th International Conference on Computer Engineering and Systems (ICCES)*, pages 657–662.

Sahlol, A. T., Kollmannsberger, P., and Ewees, A. A. (2020). Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Scientific Reports*, 10(1):2536.

Scotti, F. (2005). Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images. In *CIMSA. 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 2005.*, pages 96–101.

Zolfaghari, M. and 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.