Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hyperbolic neural networks for biologically informed classification of white blood cells and hematopoietic cell states in peripheral blood smears

Lukáš Jochymek
Technical University of Ostrava
Markéta Vašinková
marketa.vasinkova@vsb.cz
Technical University of Ostrava
Petr Gajdoš
Technical University of Ostrava

Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9930528/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Hyperbolic neural networks for biologically informed classification of white blood cells and hematopoietic cell states in peripheral blood smears

Lukáš Jochymek[1,*,+], **Markéta Vašinková**[1,*,+], **and Petr Gajdoš**[1,+]

[1]VSB - Technical University of Ostrava, Department of Computer Science, FEECS, 17. listopadu 2172/15, Ostrava, 70800, Czech Republic
[*]lukas.jochymek.st@vsb.cz, marketa.vasinkova@vsb.cz
[+]these authors contributed equally to this work

ABSTRACT

Automated white blood cell (WBC) classification in peripheral blood smears supports morphological assessment of hematological disorders, but most deep learning classifiers treat cell categories as independent labels and do not account for hematopoietic maturation. We propose a hyperbolic classification framework that combines a ResNet-18 feature extractor with a prototype-based classifier in the Poincaré ball, allowing biologically related cell states to be represented in a geometry suited to hierarchical structure. The model was evaluated on a curated dataset of 129,174 annotated single-cell images mapped to 14 blood cell classes, including leukocyte subtypes, hematopoietic precursor states, and multiple stages of granulocytic maturation. To reduce acquisition-level information leakage, we used a grouped evaluation protocol that keeps linked multi-focus acquisitions within the same subset. Across the six top-ranked grouped-split configurations, the hyperbolic model improved macro-averaged F1-score from 0.812 $\pm$ 0.005 to 0.819 $\pm$ 0.004 relative to the Euclidean baseline, while overall accuracy remained slightly lower (0.826 $\pm$ 0.002 vs 0.830 $\pm$ 0.006). Recall gains were most pronounced for promyelocytes, immature WBCs, and metamyelocytes. These results suggest that hyperbolic representation learning can improve class-balanced classification of WBCs and related hematopoietic cell states while maintaining competitive overall accuracy.

Introduction

In recent years, advances in computer vision and machine learning have substantially transformed the automated analysis of white blood cells (WBCs) in peripheral blood smears [1]. The widespread adoption of digital microscopy and slide scanning now enables high-resolution image acquisition suitable for algorithmic processing. Early approaches based on classical image processing techniques, such as color segmentation and handcrafted shape or texture analysis, were limited by strong sensitivity to morphological variability and staining conditions [2].

More recently, deep learning methods, particularly convolutional neural networks (CNNs), have markedly improved the accuracy and robustness of WBC classification. Several studies have reported performance comparable to expert morphologists; for example, Acevedo et al. [3] achieved over 95% classification accuracy using a transfer-learning CNN across eight blood cell categories. These advances have increased confidence in computer-aided diagnostic (CAD) systems as reliable tools for automated WBC differentials and for supporting early disease detection in routine hematological practice [1].

Blood cell types are not biologically independent categories but arise through hematopoietic differentiation. During this process, cells develop along specific lineages and maturation pathways, and neighboring stages within the same developmental line often share similar morphological characteristics. This is particularly relevant for immature and transitional leukocyte populations, such as myeloblasts, promyelocytes, myelocytes, metamyelocytes, band neutrophils, and mature neutrophils, where visual differences may be gradual rather than sharply separated. However, most deep learning approaches to WBC classification treat these categories as flat and independent labels. This motivates the use of representation spaces that can better reflect developmental relatedness among hematopoietic cell states while remaining applicable to the established task of automated WBC classification.

Leukemia comprises a heterogeneous group of hematologic malignancies characterized by clonal proliferation of abnormal white blood cells (WBCs). Acute leukemias, including acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), are rapidly progressing diseases originating from immature precursor cells, whereas chronic lymphocytic leukemia (CLL) follows a slower and more indolent course. AML arises from the myeloid lineage and, if untreated, can be fatal within months, while ALL originates from lymphoid precursors and predominantly affects children.

In contrast, CLL typically affects older adults and is characterized by the accumulation of mature-appearing but functionally

impaired B lymphocytes, making it the most common adult leukemia in Western countries [4]. Although its clinical course is slower, the progressive expansion of abnormal lymphocytes ultimately disrupts normal hematopoiesis and immune function [5].

CLL is diagnosed based on sustained peripheral blood monoclonal B-cell lymphocytosis according to established clinical guidelines [6]. Morphological examination of peripheral blood smears typically reveals small mature lymphocytes with dense chromatin and scant cytoplasm, frequently accompanied by smudge cells that reflect increased cellular fragility. In addition to typical small lymphocytes, larger prolymphocytes and morphologically atypical CLL variants may be present.

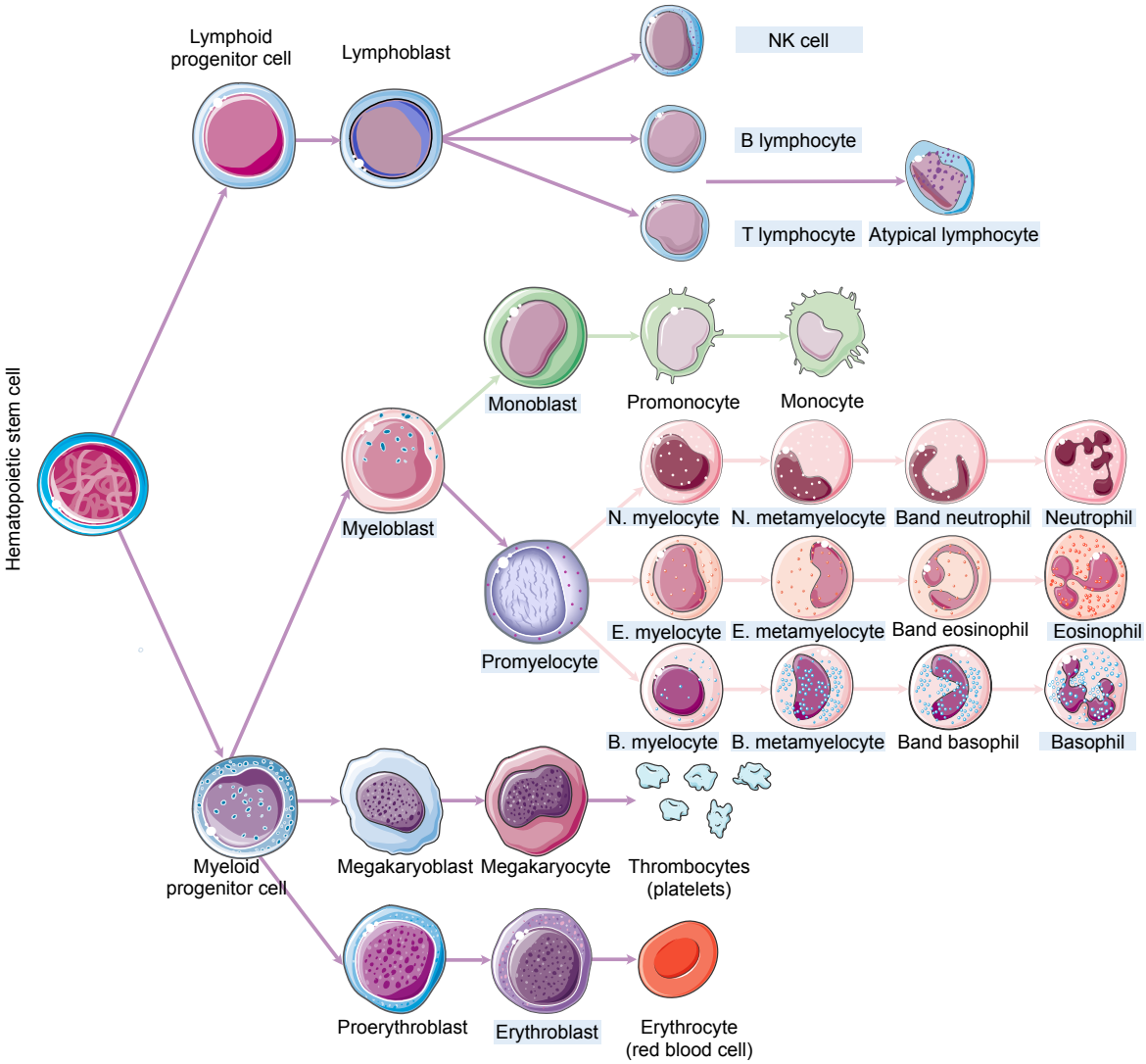


Figure 1. Conceptual illustration of the hierarchical relationships among blood cell classes. The figure highlights developmental relatedness across hematopoietic maturation stages. Blood cell types detectable in this study, including leukocytes and erythrocytes, are explicitly indicated. Due to the limited discriminative power of bright-field microscopy for certain morphologically similar cell types (e.g., different metamyelocyte subtypes), some categories are merged into a single class. In particular, metamyelocytes are not distinguished into basophilic, eosinophilic, and neutrophilic subtypes. The curated dataset comprises the following unified classes: Basophil, Blast, Eosinophil, Erythroblast, Immature WBC, Lymphocyte, Atypical lymphocyte, Metamyelocyte, Monocyte, Myeloblast, Myelocyte, Neutrophil, Band neutrophil, and Promyelocyte.

Importantly, the proportion and morphology of these cellular subtypes carry prognostic significance and are associated with disease aggressiveness and underlying molecular features [7–9]. These observations underscore the continued clinical relevance of detailed morphological assessment in CLL.

In contrast to CLL, acute leukemias are characterized by the uncontrolled proliferation of immature precursor cells (blasts), which are normally confined to the bone marrow and are virtually absent from peripheral blood. In AML, the malignant clone

arises from the myeloid lineage, whereas in ALL it originates from lymphoid progenitors. Progressive expansion of leukemic blasts disrupts normal hematopoiesis, leading to cytopenias and, in advanced stages, the dissemination of malignant cells into the peripheral blood and secondary organs [10, 11].

Acute leukemias are classified based on a combination of cytogenetic, immunophenotypic, and morphological features [12]. At the morphological level, myeloid and lymphoid blasts differ in characteristic cellular features, enabling lineage assignment and supporting diagnostic decision-making [13–15]. Although modern classification systems increasingly rely on immunophenotypic and genetic markers, morphological assessment remains clinically relevant for recognizing distinct blast populations and understanding the heterogeneity of leukemic precursor cells [16, 17].

Beyond blast evaluation, the distribution of other white blood cell subsets provides additional diagnostic and prognostic information. Alterations in the relative abundance and morphology of eosinophils, monocytes, or basophils are associated with distinct hematological disorders and disease progression [18–20]. Importantly, these patterns are not defined by a single abnormal cell type but by shifts in the composition and maturation structure of leukocyte populations. These observations illustrate that hematological diagnosis is inherently based on the relative proportions and developmental stages of leukocyte subsets rather than on isolated cell-level findings. In particular, the assessment of the granulocytic lineage relies on recognizing continuous transitions from immature precursors to fully differentiated cells, with characteristic shifts distinguishing acute leukemias from chronic or reactive conditions [21].

In this study, we aim to translate this clinical principle into an automated classification setting. Rather than treating leukocyte classes as isolated and mutually independent categories, we propose a learning framework that explicitly accounts for their developmental relatedness. To this end, we employ hyperbolic neural networks, which are well suited for modeling hierarchical structures, and evaluate the approach on a curated dataset comprising 14 blood cell classes, including leukocyte subtypes, granulocytic maturation stages, and erythroid precursors (Fig. 1).

Deep learning methods for white blood cell detection and classification

Early machine learning approaches

The first generation of automated WBC classification systems relied heavily on image preprocessing, segmentation, and handcrafted feature extraction. Typical workflows involved color normalization, nucleus and cytoplasm segmentation (e.g., via thresholding, watershed, or k-means clustering), followed by the extraction of morphological, color, and texture descriptors, such as shape factors, histogram statistics, gray-level co-occurrence matrix (GLCM) features, and local binary patterns (LBP). These features were then classified using conventional algorithms, including support vector machines (SVM), k-nearest neighbors (k-NN), random forests (RF), or multilayer perceptrons (MLP) [22, 23].

The first publicly available dataset designed for developing and evaluating algorithms for white blood cell segmentation and classification was ALL-IDB [24]. Prior to its release, no supervised image dataset existed that enabled a fair and reproducible comparison of computational methods targeting leukocyte detection and acute lymphoblastic leukemia diagnosis. ALL-IDB provided the first annotated collection of peripheral blood smear images from healthy and leukemic patients and quickly became a key benchmark for early machine-learning approaches.

Subsequent studies commonly relied on ALL-IDB to evaluate classical methods. For example, Shafique and Tehsin (2018) [25] demonstrated that handcrafted texture features combined with SVM achieved accuracies of around 90–95%, although such systems remained sensitive to staining variability and imaging conditions.

Transition to deep learning

The introduction of convolutional neural networks (CNNs) revolutionized biomedical image classification, including hematological microscopy. Early CNN-based methods such as LeNet [26], AlexNet [27], and VGG16 [28] began to replace handcrafted features by automatically learning hierarchical visual representations from raw image data [29]. Since 2017, numerous studies have emerged that use various types of convolutional neural networks for white blood cell classification. The study by Qin et al. (2018) [30] proposes a deep residual CNN architecture (Cell3Net) for fine-grained leukocyte classification, leveraging 92,000 annotated microscopic images covering 40 clinically relevant cell categories; through tailored residual blocks, extensive data augmentation, and optimized training strategies, the model achieves a top-1 accuracy of 77–78% and a top-5 accuracy of 98.7%, demonstrating that deep residual learning can handle highly detailed and morphologically subtle leukocyte distinctions far beyond traditional 5-class WBC tasks. However, performance remains limited for rare classes with scarce training samples.

The study by Hegde et al. (2019) [31] systematically compares traditional image-processing techniques with deep learning methods for classifying six white blood cell types, showing that handcrafted features combined with a neural network can achieve very high accuracy ($\approx 99.8\%$), but deep learning models, particularly CNNs trained from scratch, provide similarly strong performance ($\approx 99\%$), while eliminating the reliance on manual feature extraction.

At the same time, object detection architectures such as Faster R-CNN [32] and Mask R-CNN [33] began to unify cell localization and classification within a single framework, enabling end-to-end learning directly from whole microscopic fields

rather than pre-cropped regions [34]. This period also saw the emergence of the first transfer-learning approaches based on pretrained CNNs such as InceptionV3 [35] and ResNet50 [36], which substantially improved convergence on small biomedical datasets and increased robustness to staining variability and imaging conditions [37, 38]. For instance, Kutlu et al. (2020) [39] introduced a region-based deep learning system for simultaneous detection and classification of five white blood cell types using the Faster R-CNN framework with CNN backbones including AlexNet, VGG16, GoogLeNet, and ResNet50. By combining the Blood Cell Count and Detection Dataset BCCD [40] and LISC datasets, the authors constructed a balanced training set of 5000 images (1000 per class) and evaluated performance on 1250 test images. The best-performing configuration, Faster R-CNN with a ResNet50 feature extractor, achieved 100% WBC detection accuracy.

In recent years, Faster R-CNN has remained one of the most widely adopted architectures for leukocyte detection, owing to its reliable localization performance and strong generalization across imaging modalities. Even contemporary studies, such as the work of Zeng et al., who reported high detection accuracy across all WBC subtypes, continue to rely on Faster R-CNN as the core detection engine, further underscoring its status as a robust and well-validated baseline for cell detection tasks [41].

Modern deep learning frameworks

In the 2020s, CNN-based approaches became the standard for WBC classification. Lightweight architectures such as MobileNetV3 [42] and EfficientNet [43] have gained widespread adoption due to their favorable balance between accuracy and computational efficiency, enabling inference speeds that are suitable for deployment in clinical analyzers. MobileNetV3, in particular, has become a popular choice for compact feature extraction; for example, Rao and Rao (2023) [44] demonstrated that its partial-residual bottleneck structure can effectively capture both global and local characteristics of segmented WBC images before forwarding them to a lightweight classifier.

As detection–classification pipelines evolved, two-stage frameworks further strengthened overall robustness by explicitly separating localization from subtype recognition. These pipelines typically employ YOLO-based detectors [45] or region-proposal architectures such as R-CNN, Faster R-CNN, and Mask R-CNN, followed by a dedicated CNN classifier. Beyond classification-centric methods, Alam and Tariqul Islam (2019) [46] applied the YOLO-based framework with different classification networks as its backbones for automated full blood cell counts, using a modified BCCD dataset to simultaneously detect and enumerate red blood cells, white blood cells, and platelets. Although their system achieved sub-second inference and demonstrated good generalization across datasets, its performance on the BCCD benchmark was comparatively weak. In contrast, Xia et al. (2020) [47] applied YOLOv3, YOLOv3-SPP, and YOLOv3-tiny directly to the BCCD dataset and reported better detection outcomes. The most recent contribution in the area of two-stage approaches is the work of Wang et al. (2025) [48], who propose a hybrid framework that first detects RBCs, WBCs, and platelets using an improved YOLOv7 model enhanced with multi-head attention and the SIoU loss. In the second stage, WBC subtypes are classified using an EfficientNetV2 backbone extended with ASPP and class-balanced BCE to address data imbalance. Trained and evaluated on four public datasets (BCCD, LDWBC, LISC, and Raabin), the method achieves an mAP50 of 97.17% for multicell detection and 95.12–100% precision with 93–99% recall across WBC subtypes, outperforming single-stage detectors.

A related recent study further extends this direction by incorporating advanced optimization strategies to enhance discriminability under severe class imbalance. In leukemia-focused experiments, an EfficientNet-B3-based model combined with hybrid loss functions achieved near-perfect accuracy ($\approx 99.7\%$) across several rare leukocyte subtypes [49].

Ferdousi et al. (2024) [50] proposed a two-stage pipeline in which a super-resolution generative adversarial network (SR-GAN) is first applied to enhance fine morphological structures in low-quality blood smear images. The super-resolved images are then classified using a VGG19-based CNN to better capture subtle textural cues indicative of leukocyte subtypes. Their experiments demonstrated that SR-GAN enhancement significantly boosts discriminability, yielding higher accuracy, particularly for small or visually ambiguous classes. The study underscores the importance of image-quality restoration for improving downstream WBC classification performance.

Vision Transformers and hybrid models

Inspired by advances in natural language processing, Vision Transformers (ViTs) have recently emerged as a powerful alternative to CNNs [51]. ViTs employ self-attention to capture long-range spatial dependencies and global morphological context, which is particularly valuable for distinguishing visually similar WBC subtypes, such as eosinophils and basophils. Katar et al. (2023) [52] introduced an explainable ViT framework that generates intuitive attention maps, highlighting diagnostically relevant cellular regions. Building on this direction, hybrid CNN–ViT architectures and Swin-Transformers have achieved state-of-the-art results with improved interpretability and reduced dependence on large training sets. A recent comparative study by Fu et al. (2025) [53] further demonstrated the strength of Transformer-based backbones: across SSD, Faster R-CNNs, and YOLOv5 detectors evaluated on multiple WBC datasets, Swin-Transformer and ConvNeXt backbones consistently improved multi-scale representation and localization accuracy over traditional CNNs.

Deep learning for leukemia and cancer detection

Beyond the five canonical leukocyte types (neutrophils, eosinophils, basophils, lymphocytes, and monocytes), automated systems have been extended to detect malignant cells and diagnose hematological cancers.

Matek et al. (2019) [54] compiled an expert-annotated dataset of more than 18,000 single-cell images from peripheral blood smears and trained a convolutional neural network for leukocyte morphology classification. Their model achieved human-level performance in clinically relevant tasks, including blast-cell recognition and the identification of cell types normally present in non-pathological blood smears, marking an important milestone in computational hematopathology. Manescu et al. (2023) [55] introduced MILLIE, a weakly supervised deep-learning framework that enables reliable detection and typing of acute leukemias directly from peripheral blood smears and bone-marrow aspirates without requiring exhaustive cell-level annotations. Trained solely on patient-level diagnostic labels, the model implicitly learned disease-relevant morphological patterns and achieved high performance in distinguishing ALL, AML, and particularly acute promyelocytic leukemia (APL), with AUC values reaching up to 0.99. Park et al. (2024) [56] demonstrated that automated image-analysis approaches can be extended to blood smears containing leukemic pathology, including blasts and other immature or atypical precursor cells commonly encountered in acute leukemia. Their findings show that deep learning models are increasingly capable of coping with the pronounced morphological heterogeneity of leukemic samples, in which malignant cells coexist with normal leukocytes and often lack sharply defined morphological boundaries. Importantly, these observations highlight that blood cell populations are not naturally organized as isolated categories but rather form a continuum reflecting hematopoietic differentiation and leukemic transformation. Such a structure poses inherent limitations for models that rely on flat, purely discriminative feature spaces.

Hyperbolic deep learning

Traditional deep learning models typically operate in Euclidean vector spaces, which may be suboptimal for data with latent relational or hierarchical structures. Hyperbolic geometry provides a natural alternative because its constant negative curvature enables exponential volume growth, making it well suited for representing branching and tree-like relationships with low distortion [57, 58]. Representation-learning studies have shown that Poincaré embeddings can compactly capture both hierarchy and semantic similarity, motivating the development of hyperbolic neural networks in which neural operations and classifiers are formulated directly in hyperbolic space [59, 60]. Recent surveys further suggest that negatively curved representations may benefit computer-vision tasks involving fine-grained categorization, part-whole relationships, or latent hierarchical label structures [61]. This makes hyperbolic representation learning particularly relevant for biomedical classification tasks in which the label space reflects an implicit biological hierarchy, such as hematopoietic differentiation.

Hyperbolic convolutional neural networks

Early work on "hyperbolic deep learning" focused on extending basic neural building blocks, such as linear layers, recurrent networks, and softmax classifiers, to operate in hyperbolic manifolds, typically using the Poincaré ball model and Möbius gyrovector operations [62]. These studies established that hyperbolic networks can represent hierarchical or power-law structured data in a more compact way than their Euclidean counterparts and motivated their use in downstream tasks such as classification, retrieval, and sequence modeling [60].

The first wave of hyperbolic convolutional neural networks (hyperbolic CNNs) adopted a hybrid architecture: convolution and feature extraction are performed in Euclidean space, while the final representation and classifier are defined on a hyperbolic manifold. A representative example is the Hyperbolic Image Embeddings model of Khrulkov et al. [63], which keeps the convolutional backbone entirely Euclidean and then maps the resulting feature vectors into hyperbolic space via the exponential map. Each class is modeled by a prototype point on the hyperbolic manifold, and predictions are produced by a distance-based softmax over geodesic distances to these prototypes. On fine-grained benchmarks such as CUB-200-2011 and person re-identification datasets, hyperbolic prototypes improve few-shot and retrieval performance over purely Euclidean embeddings, while inducing a semantically meaningful geometry in which related classes cluster along tree-like directions [63].

Subsequent work investigated how curvature and manifold choice affect generalization, particularly under data scarcity and class imbalance. Gao et al. proposed a method for generating task-specific curvature in non-Euclidean spaces for few-shot learning, showing that adapting geometry to the data distribution can substantially improve performance over fixed Euclidean baselines [64]. Hyperbolic attention mechanisms, which reinterpret soft attention in a hyperbolic space, have also been shown to improve generalization in tasks where the underlying structure is hierarchical or graph-like [62]. These results, together with larger surveys [60, 61], support the view that the main advantage of hyperbolic CNNs lies in their inductive bias toward hierarchical representations and the ability to use curvature as an additional degree of modeling freedom.

At the same time, several studies revealed optimization and numerical issues in hybrid Euclidean–hyperbolic architectures. Because the Poincaré ball has a boundary where geodesic distances and gradients blow up, naive training can suffer from vanishing or exploding gradients when Euclidean features are projected with large norms. Guo et al. addressed this by introducing *clipped hyperbolic classifiers*, which simply bound the Euclidean feature norm before mapping to hyperbolic

space [65]. This stabilizes the geometry of the classifier head and allows clipped hyperbolic CNN to match or exceed Euclidean CNN performance on standard image benchmarks such as CIFAR-10, CIFAR-100, and ImageNet, while simultaneously improving adversarial robustness and out-of-distribution detection [65]. Conceptually, this line of work demonstrates that careful control of distances and curvature is crucial when coupling Euclidean backbones with hyperbolic heads.

More recently, researchers have begun to move beyond purely hybrid designs and toward *fully* hyperbolic convolutional architectures. Skliar and Weiler [66] proposed a general recipe for hyperbolic convolutional networks that systematically replaces Euclidean operations with their hyperbolic analogs, emphasizing the role of Riemannian geometry in defining convolution, aggregation, and non-linearities. Bdeir et al. introduced a fully hyperbolic CNN framework (HCNN) based on the Lorentz model of hyperbolic space, providing hyperbolic formulations of convolution, batch normalization, and multinomial logistic regression within a single end-to-end architecture [67]. Their experiments on standard vision benchmarks indicate that such fully hyperbolic CNNs can achieve competitive or superior accuracy compared to Euclidean baselines, while better preserving hierarchical relations and yielding more structured embedding spaces [61, 67].

Although most of this literature has focused on natural-image datasets, there is growing evidence that hyperbolic CNNs can be advantageous in medical imaging scenarios where labels exhibit clinical or anatomical hierarchies. Ayubcha et al. conducted one of the first systematic evaluations of hybrid Euclidean–Lorentz hyperbolic CNNs on multi-modality neuroimaging data, comparing hyperbolic ResNet architectures to standard Euclidean counterparts across several neurodegenerative and cerebrovascular disease tasks [68]. They observed comparable in-distribution accuracy but improved robustness to adversarial perturbations, more semantically organized embedding spaces, and superior zero-shot generalization to external datasets and radiologist performance [68]. These results suggest that the hierarchical inductive bias of hyperbolic geometry may be especially beneficial in biomedical applications where disease subtypes and morphological patterns naturally form tree-like structures.

The hyperbolic CNN used in this work follows the hybrid paradigm: a Euclidean ResNet-18 backbone extracts feature vectors, which are then mapped to a hyperbolic manifold where classification is performed via distance to learnable class prototypes. In contrast to earlier fixed-curvature models, the curvature magnitude c of the manifold is learned jointly with the classifier parameters, allowing the network to adapt the "amount of negative curvature" to the structure of the white blood cell label space. In addition, the classifier employs a learnable global temperature parameter that rescales the distance-based logits and thereby adjusts the sharpness of the resulting class posterior distribution. This design is closely related to prior hybrid hyperbolic classifiers [63, 65] and to curvature-learning strategies proposed in recent hyperbolic vision work [64, 67], while being tailored here to the specific morphology-driven hierarchy of leukocyte subtypes.

The overall architecture of the proposed hyperbolic prototype-based classifier and its comparison with the Euclidean CNN baseline are summarized in Fig. 2.

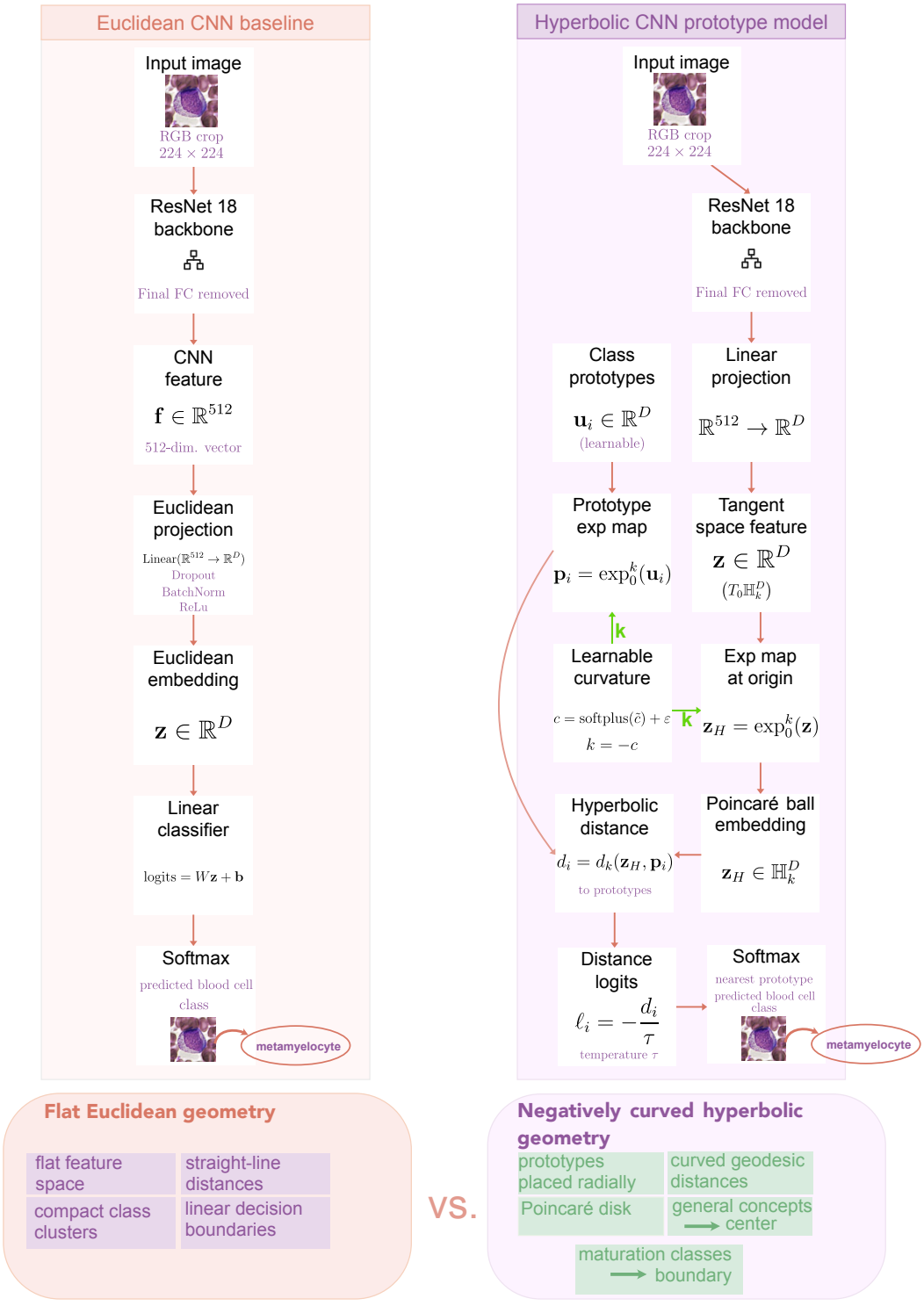


Figure 2. Overview of the Euclidean CNN baseline and the proposed hyperbolic prototype-based classifier. Both models use a ResNet-18 backbone to extract image features. In the Euclidean baseline, the projected embedding is classified using a standard linear classifier. In the hyperbolic model, the projected feature vector is mapped from the tangent space to the Poincaré ball, where classification is performed using distances to learnable class prototypes. The lower panels illustrate the conceptual difference between flat Euclidean geometry and negatively curved hyperbolic geometry for representing class relationships.

Results

We evaluated the proposed hyperbolic classification framework on a curated dataset comprising 14 blood cell classes, spanning multiple stages of granulocytic maturation, lymphoid subtypes, and erythroid precursors. Because the dataset is strongly imbalanced, model performance was assessed primarily using the macro-averaged F1-score, which gives equal importance to all classes regardless of their frequency.

During preliminary analyzes, we identified that a random image-level split could introduce acquisition-level information leakage in the multi-focus WBC subset, where different focal-plane images of the same physical cell may be assigned to both training and test sets. Such a split may overestimate performance because the test set can contain images that are highly similar to samples already observed during training, differing mainly in focal plane rather than biological content. Therefore, all results reported below are based on an acquisition-aware grouped split that keeps linked multi-focus acquisitions within the same subset, providing a more conservative and reliable estimate of generalization to unseen cells.

Table 1 summarizes the Euclidean baseline results, and Table 2 reports the proposed hyperbolic neural network results, both evaluated under this grouped evaluation protocol.

		Validation		Test	
Acc.	F1 macro	Sens. (w)	Spec. (w)	Sens. (w)	Spec. (w)
0.829	**0.819**	**0.831**	**0.959**	**0.829**	**0.958**
0.841	0.818	0.835	0.966	0.841	0.969
0.825	0.812	0.828	0.967	0.825	0.966
0.829	0.810	0.833	0.956	0.829	0.956
0.828	0.809	0.828	0.967	0.828	0.968
0.829	0.805	0.832	0.965	0.829	0.964
0.816	0.803	0.823	0.959	0.816	0.958
0.812	0.801	0.813	0.967	0.812	0.968
0.820	0.799	0.816	0.958	0.820	0.959
0.823	0.794	0.830	0.958	0.823	0.957
0.820	0.792	0.816	0.956	0.820	0.957
0.819	0.789	0.822	0.959	0.819	0.958
0.802	0.787	0.799	0.967	0.802	0.969
0.815	0.782	0.812	0.960	0.815	0.962
0.804	0.766	0.805	0.957	0.804	0.956
0.791	0.756	0.790	0.963	0.791	0.963
0.796	0.741	0.792	0.959	0.796	0.961
0.765	0.725	0.761	0.954	0.765	0.954
0.791	0.724	0.789	0.960	0.791	0.961
0.768	0.707	0.763	0.952	0.768	0.954

Table 1. Validation and test performance metrics of the Euclidean baseline classifier under the grouped evaluation protocol. The table reports all 20 grouped-split configurations, ordered by test macro-F1. F1 macro is reported as the primary evaluation metric. Sensitivity (Sens.) and specificity (Spec.) are reported as supplementary weighted metrics to account for class imbalance.

Across the six best-performing grouped reruns, the tuned hyperbolic model achieved a higher mean macro-F1 than the Euclidean baseline, indicating improved class-balanced performance under the grouped evaluation protocol. The Euclidean model retained a small advantage in mean accuracy and weighted sensitivity, while weighted specificity remained very similar between the two models. This suggests that the benefit of the hyperbolic representation is most evident when performance is evaluated in a class-balanced manner rather than by overall accuracy alone. To assess whether this pattern was limited to the selected best-performing configurations, we also compared both models across all 20 grouped-split configurations. In this broader comparison, the hyperbolic model achieved higher average macro-F1, accuracy, and weighted sensitivity, while also showing substantially lower variability across runs. The difference was most pronounced for macro-F1, where the hyperbolic model reached 0.812 ± 0.006 compared with 0.782 ± 0.034 for the Euclidean baseline. Accuracy and weighted sensitivity showed the same trend, increasing from 0.811 ± 0.020 for the Euclidean model to 0.825 ± 0.004 for the hyperbolic model. Weighted specificity remained high and comparable for both classifiers. Thus, the hyperbolic model did not consistently outperform the Euclidean baseline in all metrics, but it showed a clearer advantage in macro-F1 and produced more consistent results across the evaluated configurations.

		Validation		Test	
Acc.	F1 macro	Sens. (w)	Spec. (w)	Sens. (w)	Spec. (w)
0.827	**0.824**	**0.834**	**0.963**	**0.827**	**0.962**
0.828	0.821	0.821	0.962	0.828	0.964
0.826	0.820	0.829	0.960	0.826	0.960
0.824	0.819	0.832	0.962	0.824	0.962
0.828	0.815	0.829	0.961	0.828	0.963
0.824	0.813	0.833	0.962	0.824	0.960
0.823	0.813	0.822	0.963	0.823	0.963
0.829	0.813	0.826	0.961	0.829	0.963
0.832	0.813	0.822	0.962	0.832	0.964
0.823	0.812	0.831	0.964	0.823	0.963
0.831	0.812	0.829	0.965	0.831	0.965
0.826	0.810	0.820	0.962	0.826	0.964
0.824	0.809	0.817	0.961	0.824	0.963
0.818	0.809	0.817	0.967	0.818	0.968
0.819	0.808	0.827	0.966	0.819	0.964
0.826	0.807	0.833	0.963	0.826	0.962
0.824	0.807	0.820	0.960	0.824	0.961
0.819	0.806	0.823	0.963	0.819	0.962
0.815	0.803	0.825	0.958	0.815	0.956
0.827	0.803	0.822	0.960	0.827	0.962

Table 2. Validation and test performance metrics of the proposed hyperbolic neural network under the grouped evaluation protocol. The table reports all 20 grouped-split configurations, ordered by test macro-F1. F1 macro is reported as the primary evaluation metric. Sensitivity (Sens.) and specificity (Spec.) are reported as supplementary weighted metrics to account for class imbalance.

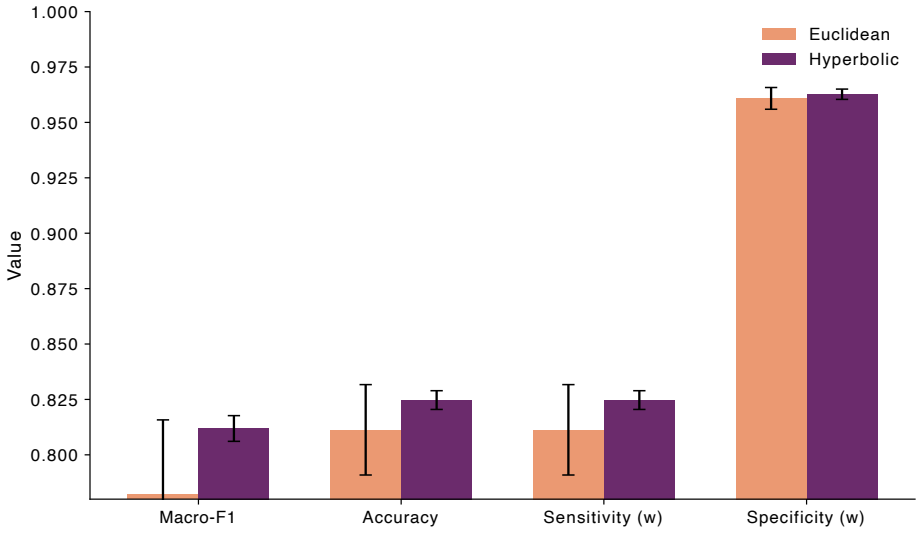


Figure 3. Mean ± standard deviation of test-set performance metrics for the Euclidean and hyperbolic classifiers, computed over all 20 grouped-split configurations. The comparison summarizes overall performance stability across the evaluated configurations. Macro-F1 was used as the primary metric because of the pronounced class imbalance, while accuracy, weighted sensitivity, and weighted specificity are reported as supplementary metrics.

The TOP-6 comparison in Table 3 provides a more focused view of the best-performing configurations of each model. Under this selection, the tuned hyperbolic model still provides a stronger macro-F1 summary, whereas the Euclidean baseline maintains a modest advantage in test accuracy and weighted sensitivity. In this single-label multi-class setting, weighted sensitivity closely tracks overall accuracy, so the main gain of the tuned hyperbolic setup lies in improved class-balanced performance rather than in overall accuracy alone.

Metric	Euclidean (mean ± std)	Hyperbolic (mean ± std)	Δ mean (H−E)
Val. Sens. (w)	0.831 ± 0.003	0.830 ± 0.005	−0.002
Val. Spec. (w)	0.963 ± 0.005	0.962 ± 0.002	−0.001
Test Acc.	0.830 ± 0.006	0.826 ± 0.002	−0.004
Test Macro-F1	**0.812 ± 0.005**	**0.819 ± 0.004**	**0.007**
Test Sens. (w)	0.830 ± 0.006	0.826 ± 0.002	−0.004
Test Spec. (w)	0.963 ± 0.005	0.962 ± 0.001	−0.002
Test TN (w)	16013.097 ± 64.637	15992.013 ± 16.345	−21.084
Test TP (total)	16088.000 ± 109.278	16006.167 ± 32.314	−81.833
Test FP (total)	3287.000 ± 109.278	3368.833 ± 32.314	81.833
Test FN (total)	3287.000 ± 109.278	3368.833 ± 32.314	81.833
Best epoch	15.500 ± 3.209	23.167 ± 1.472	7.667

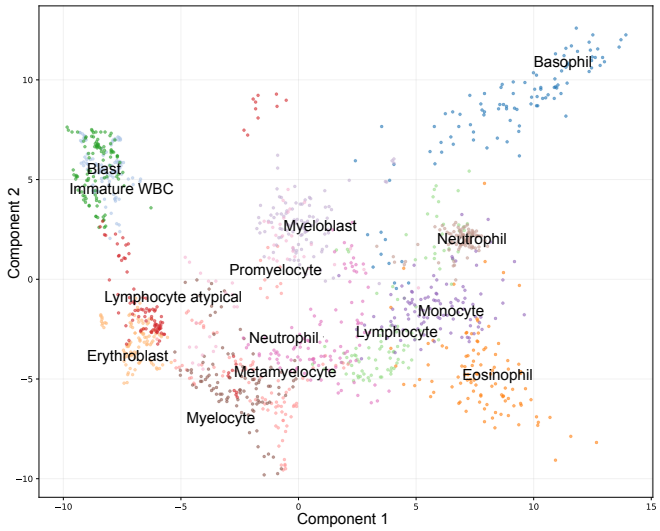
Table 3. Comparison between Euclidean and hyperbolic classifiers using the six best-performing grouped-split configurations (TOP-6) of each model, selected by test macro-F1. Reported values denote mean ± standard deviation. Weighted true negatives (TN) are included to complement sensitivity and specificity under class imbalance. In multi-class single-label classification, total false positives and false negatives are identical and reflect the overall number of misclassified samples.

Class	Recall (E)	Recall (H)	F1 (E)	F1 (H)
Blast	0.571	0.514	0.592	0.572
Immature WBC	0.813	0.839	0.791	0.797
Promyelocyte	0.187	0.371	0.243	0.413
Myeloblast	0.936	0.937	0.903	0.916
Myelocyte	0.853	0.828	0.850	0.829
Metamyelocyte	0.553	0.578	0.563	0.556

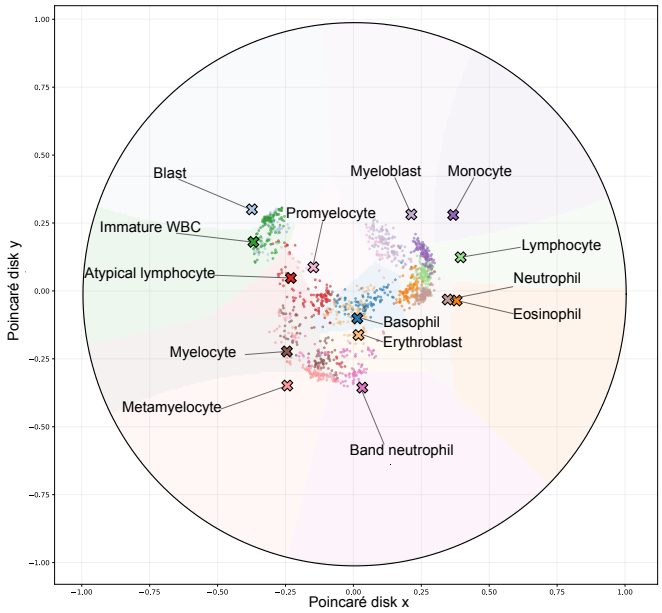
Table 4. Clinically oriented per-class test performance for selected precursor and transitional cell populations. Reported values are mean recall and mean F1 across the six best-performing grouped-split configurations of each model.

The clinically oriented view in Table 4 clarifies where the hyperbolic representation helps the most. The largest gain was observed for promyelocytes, where the hyperbolic model nearly doubled the mean recall and substantially improved the F1 score. More modest recall gains were also observed for immature WBCs and metamyelocytes, while myeloblast performance remained essentially comparable, with a slight F1 advantage for the hyperbolic model. In contrast, the Euclidean baseline remained stronger for blasts and myelocytes. This pattern suggests that the main benefit of the hyperbolic representation lies in better balancing performance across morphologically adjacent precursor states rather than uniformly improving every class.

To further examine how both models organize leukocyte classes in the learned representation space, we visualized the test-set embeddings of the best Euclidean and hyperbolic classifiers (Fig. 4). For the Euclidean baseline, the high-dimensional test embeddings were projected into two dimensions using principal component analysis (PCA). For the hyperbolic model, the embeddings were visualized on the Poincaré disk together with the learned class prototypes.



(a) Test-set embeddings of the best Euclidean model projected into two dimensions using PCA.



(b) Test-set embeddings of the best hyperbolic model visualized on the Poincaré disk. Cross markers indicate learned class prototypes, and shaded regions show the corresponding prototype-based decision regions.

Figure 4. Qualitative visualization of the learned representation spaces for the Euclidean and hyperbolic classifiers. In the Euclidean PCA projection, several morphologically distinct classes form separate regions, whereas precursor and transitional leukocyte populations remain partially intermixed. In the hyperbolic representation, the learned class prototypes organize cell classes in a more structured geometry, with biologically related populations positioned in neighbouring regions of the disk. These visualizations are intended as qualitative support for interpreting the learned embeddings and should not be read as a direct quantitative comparison, because PCA projection and Poincaré disk visualization preserve different geometric properties.

The embedding visualizations provide additional qualitative support for the quantitative findings. While the Euclidean PCA projection reflects the separation of several visually distinct classes, precursor and transitional populations remain partially overlapping in the projected space. In contrast, the hyperbolic representation suggests a more structured organization of developmentally related cell types, consistent with the intended hierarchical inductive bias of the proposed model.

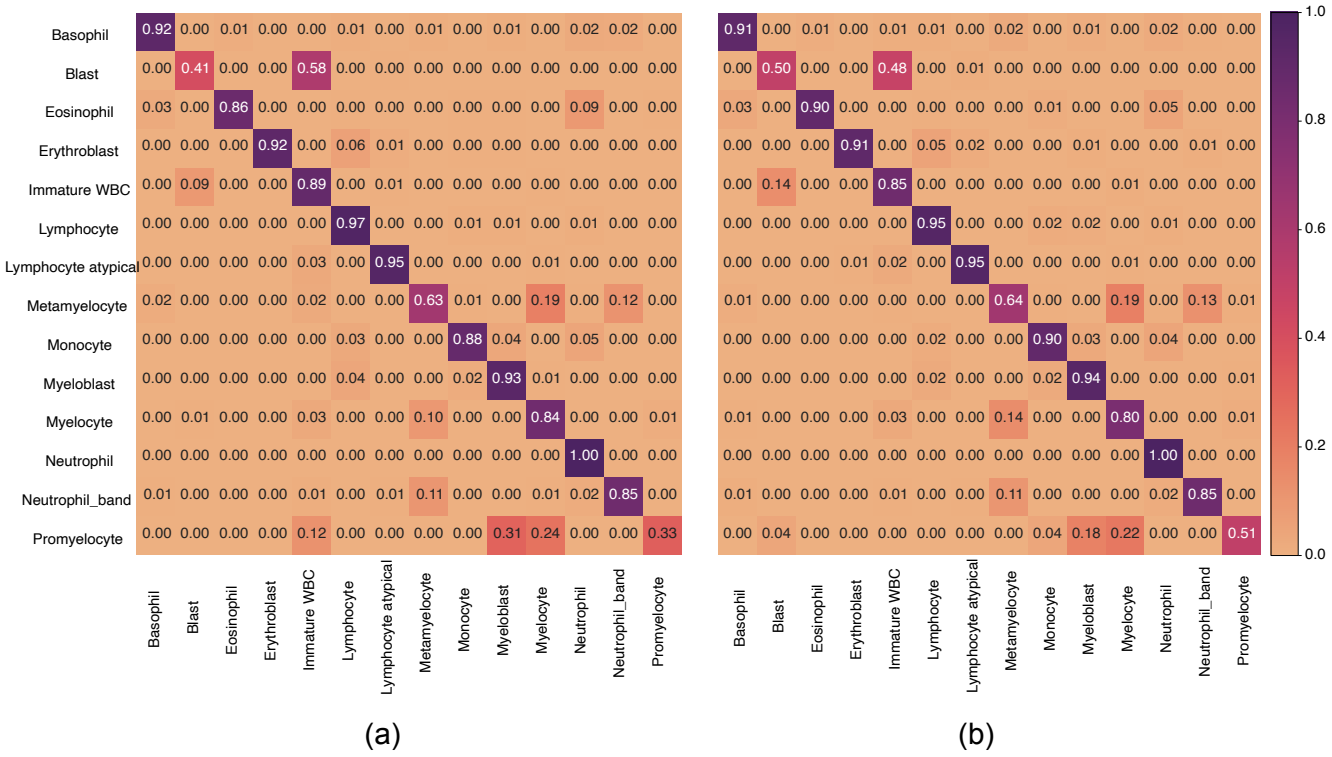


Figure 5. Row-normalized confusion matrices for the Euclidean (a) and hyperbolic (b) classifiers on the test set. Rows correspond to ground-truth classes and columns to predicted classes, with values indicating the proportion of each true class assigned to a given predicted class. The matrices provide a class-level view of the error structure, particularly among morphologically similar leukocyte subtypes.

The confusion matrices in Fig. 5 provide a class-level view of the error structure for representative Euclidean and hyperbolic classifiers. Misclassifications are not random but are concentrated between biologically and morphologically related classes. The most prominent errors occur between blast-like and immature WBC categories and among adjacent stages of granulocytic maturation, including myelocytes, metamyelocytes, band neutrophils, and neutrophils. These visualized single-model results should be interpreted together with the averaged TOP-6 results in Table 4, which provide the main quantitative summary for selected precursor and transitional populations.

Discussion

Because the three data sources differ in label granularity and terminology, we adopted a unified class taxonomy comprising 14 cell types and mapped each source-specific label to this common set. In particular, we preserved a specific myeloblast class (available in the Matek dataset and consistent with established myeloid blast morphology), while also retaining a broader blast category used by morphology-only sources as an umbrella label for immature blast-like cells without reliable lineage confirmation.

Importantly, the unified taxonomy explicitly captures the granulocytic maturation continuum, which gives rise to neutrophilic, eosinophilic, and basophilic lineages. This continuum spans from myeloblasts through promyelocytes, myelocytes, and metamyelocytes to band and fully mature granulocytes. The relative distribution of cells along this developmental axis is of direct clinical relevance, as it provides essential diagnostic and prognostic information in hematological malignancies. For example, an accumulation of early precursors and blasts is characteristic of acute myeloid leukemia, whereas a left shift with increased promyelocytes, myelocytes, and metamyelocytes is commonly observed in myeloproliferative and myelodysplastic disorders. In addition, disproportionate expansion of eosinophilic or basophilic granulocytes may point to specific cytogenetic

subtypes of myeloid neoplasms. Consequently, accurate quantification of individual maturation stages across all granulocytic branches is a cornerstone of morphological leukemia assessment and cannot be reduced to a simple binary distinction between blasts and mature cells. This clinically grounded, hierarchical organization of myeloid differentiation naturally results in a nested label structure (blast-like cells → lineage-specific precursors → mature granulocytes), which closely aligns with our motivation for hyperbolic representation learning.

From the multi-focus WBC dataset, we retained a subset of classes that are morphologically consistent with our unified taxonomy and informative for modeling leukocyte maturation and pathological states. The following categories were preserved without modification: immature WBC, band neutrophil, basophil, blast, metamyelocyte, myelocyte, and promyelocyte. The dataset's abnormal lymphocyte category was not treated as a separate class and was merged with the atypical lymphocyte class used in Matek's dataset, as these terms represent equivalent nomenclature for the same morphological category across different data sources. In addition, nucleated red blood cells (nRBCs) were mapped to the erythroblast class, reflecting consistent labeling of nucleated erythroid precursors across datasets. This harmonization strategy preserves the detailed representation of immature myeloid stages and blast-like cells provided by the multi-focus acquisition while ensuring consistency with the class definitions adopted from the AML and Raabin-WBC datasets. Moreover, it maintains a hierarchical organization of precursor cells (e.g., immature WBC → blast → myeloblast), which is particularly well suited for evaluating geometry-aware models such as hyperbolic neural networks.

Methods

In this study, we use the term white blood cell (WBC) classification when referring to the established computer-vision task and the terminology commonly used in previous deep learning studies. However, the harmonized taxonomy used in our experiments also includes hematopoietic precursor states and erythroid precursors, such as erythroblasts. Therefore, when describing the final dataset and class taxonomy, we use the broader terms blood cell classes or hematopoietic cell states. This distinction reflects the difference between the conventional task terminology and the biological scope of the curated classification problem.

Datasets and benchmarking

The dataset used in this study was curated from three publicly available sources to ensure variability in staining protocols, imaging conditions, and annotation granularity. Specifically, we integrated data from (i) the AML-Cytomorphology dataset by Matek et al., (ii) the Raabin-WBC dataset, and (iii) a multi-focus white blood cell (WBC) dataset. All datasets consist of cropped single-cell images acquired from peripheral blood smears and annotated by expert hematopathologists or trained specialists. Due to differences in labeling schemes and class granularity across the data sources, all samples were mapped to a unified taxonomy comprising 14 blood cell classes, including leukocyte subtypes and erythroid precursors. This harmonization enables joint training and benchmarking across heterogeneous datasets while preserving biologically meaningful distinctions between major leukocyte subtypes and maturation stages.

Train/validation/test partitions were generated using an acquisition-aware grouped split with a 70/15/15 ratio. In particular, all ten focal slices belonging to the same multi-focus z-stack were assigned to the same split, and repeated image variants sharing the same recoverable acquisition identifier in the merged folder structure were likewise kept together. This protocol removes the main acquisition-level leakage path in the combined benchmark. However, because patient identifiers are not preserved consistently for all source datasets in the merged ImageFolder release, the final split protocol should be described as acquisition-aware rather than fully patient-level across all merged sources.

AML-Cytomorphology dataset

The AML-Cytomorphology dataset by Matek et al. [69] provides single-cell image patches extracted from peripheral blood smears of patients with acute myeloid leukemia and non-malignant controls. The dataset includes detailed morphological annotations covering multiple stages of myeloid maturation and blast subtypes. From this dataset, we selected leukocyte classes that are morphologically well-defined and compatible with cross-dataset harmonization. Specifically, basophils, eosinophils, erythroblasts, typical and atypical lymphocytes, monocytes, myelocytes, metamyelocytes, band and segmented neutrophils, promyelocytes, and myeloblasts were retained. Typical lymphocytes were mapped to a unified lymphocyte class, while atypical lymphocytes were preserved as a separate category. Monoblasts were not treated as an independent class and were merged into a broader blast category, reflecting the limited reliability of morphology alone for distinguishing monoblasts from other blast-like cells without immunophenotypic confirmation. In contrast, myeloblasts were retained as a distinct class due to their characteristic and consistently annotated morphological features within the AML dataset.

Raabin-WBC dataset

The Raabin-WBC dataset by Kouzehkanan et al. [70] is a large-scale collection of peripheral blood smear images designed for automated leukocyte analysis. It provides expert annotations and captures substantial variability in staining and acquisition conditions. In this study, only the original five-class WBC subset was used, comprising basophils, eosinophils, lymphocytes,

monocytes, and neutrophils. No further subdivision of lymphocytes was performed, and all lymphoid cells were treated as a single lymphocyte class, following the original dataset definition. Annotations related to artifacts, smudge cells, or extended subclasses available in later versions of the dataset were not considered in order to maintain consistency with routine five-part differential counts.

Multi-focus WBC dataset

The multi-focus WBC dataset by Park et al. [56] provides z-stack image acquisitions in which each leukocyte is captured across multiple focal planes. Each sample consists of ten focal slices of size 200×200 pixels, acquired using a $50\times$ objective lens with a 400 nm step size. The dataset comprises 25,773 multi-focus image stacks collected from 72 patients and includes both normal and abnormal leukocyte categories. From this dataset, we retained classes that are morphologically consistent with the unified taxonomy and informative for modeling leukocyte maturation. Specifically, immature WBCs, blasts, basophils, band neutrophils, metamyelocytes, myelocytes, and promyelocytes were preserved. Abnormal lymphocytes were merged with the atypical lymphocyte class defined in the AML dataset, and nucleated red blood cells were mapped to the erythroblast class. This conservative class mapping reflects the morphology-based nature of the annotations and avoids introducing unreliable lineage distinctions.

Unified class taxonomy

After harmonization, the final dataset comprises 14 blood cell classes: Basophil (1,651), Blast (18,646), Eosinophil (1,490), Erythroblast (1,508), Immature WBC (35,190), Lymphocyte (7,546), Atypical lymphocyte (10,461), Metamyelocyte (5,075), Monocyte (2,584), Myeloblast (3,268), Myelocyte (10,912), Neutrophil (19,346), Band neutrophil (11,169), and Promyelocyte (328). This taxonomy explicitly represents multiple stages of granulocytic maturation and preserves developmentally related class groupings, which are subsequently exploited by the proposed hyperbolic representation learning framework.

Problem formulation

We formulate white blood cell recognition as a multi-class, single-label image classification task. Given an input image of a single blood cell, the objective is to assign exactly one class label from a predefined set of $K = 14$ blood cell classes. The class distribution is highly imbalanced, reflecting both the natural prevalence of different leukocyte types and the heterogeneous composition of the integrated datasets. Importantly, the target classes are not assumed to be independent. Instead, they exhibit structured relationships arising from hematopoietic maturation, particularly within the granulocytic lineage. Several classes correspond to successive stages along a developmental continuum (e.g., myeloblast $\rightarrow$ promyelocyte $\rightarrow$ myelocyte $\rightarrow$ metamyelocyte $\rightarrow$ band neutrophil $\rightarrow$ neutrophil), resulting in a nested label structure with varying degrees of morphological similarity.

From a modeling perspective, this implies that misclassifications between developmentally adjacent classes are more likely than confusions between distant or unrelated categories. Consequently, the classification problem is characterized not only by inter-class imbalance, but also by non-uniform semantic distances between labels. These properties motivate the use of representation spaces that can explicitly encode hierarchical and tree-like relationships among classes, rather than treating all class labels as equidistant.

Hyperbolic CNN with prototype-based classification

The proposed method combines convolutional feature extraction with hyperbolic geometry to perform distance-based classification of white blood cell images. The overall architecture consists of a convolutional neural network (CNN) backbone followed by a hyperbolic prototype-based classification head, following the hybrid Euclidean–hyperbolic paradigm used in prior hyperbolic vision models [63, 65]. Input images are first processed by a ResNet-18 backbone [36], which serves as a feature extractor. The final fully connected layer of the network is removed, and the resulting representation is passed through a learnable linear projection layer to obtain a fixed-dimensional Euclidean embedding

$$\mathbf{z} = f_\theta(\mathbf{x}) \in \mathbb{R}^D, \tag{1}$$

where $\mathbf{x}$ denotes the input image and f_θ denotes the trainable backbone together with the projection layer. The backbone is fine-tuned during training to adapt feature representations to the target domain.

To better capture hierarchical and fine-grained relationships between cell classes, the Euclidean feature embeddings are projected into a hyperbolic space modeled as a Poincaré ball [58, 62]. The geometry of this space is controlled by a learnable positive curvature magnitude c, implemented through a constrained reparameterization of an unconstrained scalar $\tilde{c}$:

$$c = \mathrm{clip}\big(\mathrm{softplus}(\tilde{c}) + \varepsilon_c,\, c_{\min},\, c_{\max}\big), \qquad k = -c. \tag{2}$$

Here, $k < 0$ denotes the sectional curvature used by the hyperbolic operations. The Euclidean embedding is then mapped to the hyperbolic manifold by the exponential map at the origin,

$$\mathbf{z}_{\mathbb{H}} = \exp_{\mathbf{0}}^{k}(\mathbf{z}). \tag{3}$$

Using a learnable curvature follows recent work showing that adapting the geometry to the data can improve representation quality and classification performance in hyperbolic models [64, 67].

Classification is performed using a prototype-based approach in hyperbolic space, similar to prior hyperbolic image embedding methods [63]. Rather than optimizing the class prototypes directly on the manifold, the model learns tangent-space parameters $\mathbf{u}_i \in \mathbb{R}^D$ for each class i, which are mapped to the Poincaré ball using the same exponential map:

$$\mathbf{p}_i = \exp_{\mathbf{0}}^{k}(\mathbf{u}_i). \tag{4}$$

The class-specific hyperbolic distances are then computed as

$$d_i = d_k(\mathbf{z}_{\mathbb{H}}, \mathbf{p}_i). \tag{5}$$

These distances are converted into logits using a learnable global temperature parameter τ, which is likewise constrained to remain positive:

$$\tau = \mathrm{clip}\big(\mathrm{softplus}(\tilde{\tau}) + \varepsilon_\tau,\ \tau_{\min},\ \tau_{\max}\big). \tag{6}$$

The logits and posterior class probabilities are therefore given by

$$\ell_i = -\frac{d_i}{\tau}, \qquad P(y = i \mid \mathbf{x}) = \frac{\exp(\ell_i)}{\sum_{j=1}^{K} \exp(\ell_j)}. \tag{7}$$

This distance-based softmax formulation is consistent with prior hyperbolic prototype classifiers, while the learnable temperature controls the sharpness of the posterior distribution [63, 65]: smaller values induce more decisive class separation, whereas larger values yield smoother posterior distributions.

The entire model is trained end-to-end using the cross-entropy loss. Optimization jointly updates the CNN backbone parameters, the projection layer, the tangent-space prototype parameters, the raw curvature parameter, and the raw temperature parameter. Different learning rates are employed for the backbone and hyperbolic components to ensure stable convergence.

Training configurations

The Euclidean baseline and the hyperbolic classifier were evaluated across 20 grouped-split configurations. The corresponding hyperparameter settings are provided in Supplementary Tables S1 and S2 online. These Supplementary Tables correspond directly to Tables 1 and 2, respectively, and follow the same ordering by test macro-F1.

Implementation details and computational environment

The implementation was carried out in Python 3.13.3 using the deep learning framework and package versions specified in the accompanying reproducibility package. The computational setup used for training included 20 Intel(R) Xeon(R) Gold 6154 CPU cores (@ 3.00 GHz), 2 NVIDIA Tesla V100 GPUs with 32 GB memory, and 7.3 TB of RAM, running Red Hat Enterprise Linux. The complete source code, configuration files, environment specification, trained checkpoints, and scripts for reproducing the reported tables and figures are provided in the accompanying reproducibility package.

Model evaluation and performance metrics

Because the dataset is strongly imbalanced, macro-averaged F1-score was used as the primary evaluation metric. Accuracy, sensitivity, and specificity were reported as supplementary metrics. Model performance was assessed using standard classification metrics: precision, recall (sensitivity), specificity, F1-score, and accuracy. These measures are widely used in the evaluation of machine learning classifiers, including medical image analysis systems, as they quantify different aspects of predictive performance based on the confusion-matrix counts of True Positives (TP), False Negatives (FN), True Negatives (TN), and False Positives (FP) [71, 72]. In multi-class single-label classification, each misclassification contributes one false positive and one false negative across the class set; therefore, the reported total FP and FN values are identical and reflect the overall number of errors.

Ethics statement

This study used only publicly available, de-identified image datasets. No new human participants, biological samples, or patient-identifiable data were collected for this work. Therefore, no additional ethical approval or informed consent was required for the present secondary analysis.

References

1. Choudhary, S. *et al.* Advancing blood cell detection and classification: performance evaluation of modern deep learning models. *BMC Med. Informatics Decis. Mak.* **25**, 207 (2025).

2. Aksoy, A. An innovative hybrid model for automatic detection of white blood cells in clinical laboratories. *Diagnostics* **14**, 2093 (2024).

3. Acevedo, A., Alférez, S., Merino, A., Puigví, L. & Rodellar, J. Recognition of peripheral blood cell images using convolutional neural networks. *Comput. methods programs biomedicine* **180**, 105020 (2019).

4. Scarfò, L., Ferreri, A. J. & Ghia, P. Chronic lymphocytic leukaemia. *Critical reviews oncology/hematology* **104**, 169–182 (2016).

5. O'Reilly, A., Murphy, J., Rawe, S. & Garvey, M. Chronic lymphocytic leukemia: a review of front-line treatment options, with a focus on elderly cll patients. *Clin. Lymphoma Myeloma Leuk.* **18**, 249–256 (2018).

6. Hallek, M. *et al.* Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the international workshop on chronic lymphocytic leukemia updating the national cancer institute–working group 1996 guidelines. *Blood, The J. Am. Soc. Hematol.* **111**, 5446–5456 (2008).

7. Oscier, D. *et al.* The morphology of cll revisited: the clinical significance of prolymphocytes and correlations with prognostic/molecular markers in the lrf cll4 trial. *Br. journal haematology* **174**, 767–775 (2016).

8. Hallek, M. Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. *Am. journal hematology* **94**, 1266–1287 (2019).

9. Matutes, E. & Polliack, A. Morphological and immunophenotypic features of chronic lymphocytic leukemia. *Rev. clinical experimental hematology* **4**, 22–47 (2000).

10. Appelbaum, F. R. 95 - acute leukemias in adults. In Niederhuber, J. E., Armitage, J. O., Kastan, M. B., Doroshow, J. H. & Tepper, J. E. (eds.) *Abeloff's Clinical Oncology (Sixth Edition)*, 1783–1797.e1, DOI: https://doi.org/10.1016/B978-0-323-47674-4.00095-5 (Elsevier, Philadelphia, 2020), sixth edition edn.

11. Park, H. S. What is new in acute myeloid leukemia classification? *Blood research* **59**, 15 (2024).

12. Asou, N. The role of a runt domain transcription factor aml1/runx1 in leukemogenesis and its clinical implications. *Critical reviews oncology/hematology* **45**, 129–150 (2003).

13. Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T. Chapter 18 - acute myeloid leukemia—overview. In Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T. (eds.) *Atlas of Hematopathology (Second Edition)*, 293–302, DOI: https://doi.org/10.1016/B978-0-12-809843-1.00018-8 (Academic Press, 2018), second edition edn.

14. Chiaretti, S., Zini, G. & Bassan, R. Diagnosis and subclassification of acute lymphoblastic leukemia. *Mediterr. journal hematology infectious diseases* **6**, e2014073 (2014).

15. Sekar, M. D., Raj, M. & Manivannan, P. Role of morphology in the diagnosis of acute leukemias: Systematic review. *Indian J. Med. Paediatr. Oncol.* **44**, 464–473 (2023).

16. Swerdlow, S. H. *et al. WHO classification of tumours of haematopoietic and lymphoid tissues*, vol. 2 (International agency for research on cancer Lyon, France, 2008).

17. Arber, D. A. *et al.* International consensus classification of myeloid neoplasms and acute leukemias: integrating morphologic, clinical, and genomic data. *Blood, The J. Am. Soc. Hematol.* **140**, 1200–1228 (2022).

18. Naymagon, L., Marcellino, B. & Mascarenhas, J. Eosinophilia in acute myeloid leukemia: overlooked and underexamined. *Blood Rev.* **36**, 23–31 (2019).

19. Addinsell, H. M. *et al.* Multi-omic analysis of chronic myelomonocytic leukemia monocytes reveals metabolic and immune dysregulation leading to altered macrophage polarization. *Leukemia* **39**, 770–774 (2025).

20. Valent, P., Horny, H.-P. & Arock, M. The underestimated role of basophils in ph+ chronic myeloid leukaemia. *Eur. journal clinical investigation* **48**, e13000 (2018).

21. Arber, D. A. & Orazi, A. Update on the pathologic diagnosis of chronic myelomonocytic leukemia. *Mod. Pathol.* **32**, 732–740 (2019).

22. Rezatofighi, S. H. & Soltanian-Zadeh, H. Automatic recognition of five types of white blood cells in peripheral blood. *Comput. Med. Imaging Graph.* **35**, 333–343 (2011).

23. Putzu, L., Caocci, G. & Di Ruberto, C. Leucocyte classification for leukaemia detection using image processing techniques. *Artif. intelligence medicine* **62**, 179–191 (2014).

24. Labati, R. D., Piuri, V. & Scotti, F. All-idb: The acute lymphoblastic leukemia image database for image processing. In *2011 18th IEEE international conference on image processing*, 2045–2048 (IEEE, 2011).

25. Shafique, S. & Tehsin, S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technol. cancer research & treatment* **17**, 1533033818802789 (2018).

26. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (2002).

27. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. neural information processing systems* **25** (2012).

28. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

29. Rehman, A. *et al.* Classification of acute lymphoblastic leukemia using deep learning. *Microsc. Res. Tech.* **81**, 1310–1317 (2018).

30. Qin, F. *et al.* Fine-grained leukocyte classification with deep residual learning for microscopic images. *Comput. methods programs biomedicine* **162**, 243–252 (2018).

31. Hegde, R. B., Prasad, K., Hebbar, H. & Singh, B. M. K. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybern. Biomed. Eng.* **39**, 382–392 (2019).

32. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. neural information processing systems* **28** (2015).

33. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969 (2017).

34. Dhieb, N., Ghazzai, H., Besbes, H. & Massoud, Y. An automated blood cells counting and classification framework using mask r-cnn deep learning model. In *2019 31st international conference on microelectronics (ICM)*, 300–303 (IEEE, 2019).

35. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

36. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).

37. Habibzadeh, M., Jannesari, M., Rezaei, Z., Baharvand, H. & Totonchi, M. Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception. In *Tenth international conference on machine vision (ICMV 2017)*, vol. 10696, 274–281 (SPIE, 2018).

38. Chen, H. *et al.* Accurate classification of white blood cells by coupling pre-trained resnet and densenet with scam mechanism. *BMC bioinformatics* **23**, 282 (2022).

39. Kutlu, H., Avci, E. & Özyurt, F. White blood cells detection and classification based on regional convolutional neural networks. *Med. hypotheses* **135**, 109472 (2020).

40. Mooney, P. Bccd: Blood cell count and detection dataset. https://www.kaggle.com/datasets/paultimothymooney/blood-cells (2018). Kaggle dataset.

41. Zeng, F. *et al.* Rapid detection of white blood cells using hyperspectral microscopic imaging system combined with multi-data faster rcnn. *Sensors Actuators B: Chem.* **389**, 133865 (2023).

42. Howard, A. *et al.* Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1314–1324 (2019).

43. Tan, M. & Le, Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In Chaudhuri, K. & Salakhutdinov, R. (eds.) *Proceedings of the 36th International Conference on Machine Learning*, vol. 97 of *Proceedings of Machine Learning Research*, 6105–6114 (PMLR, 2019).

44. Rao, B. S. S. & Rao, B. S. An effective wbc segmentation and classification using mobilenetv3–shufflenetv2 based deep learning framework. *IEEE Access* **11**, 27739–27748 (2023).

45. Mao, Y. *et al.* Dws-yolo: A lightweight detector for blood cell detection. *Appl. Artif. Intell.* **38**, 2318673 (2024).

46. Alam, M. M. & Islam, M. T. Machine learning approach of automatic identification and counting of blood cells. *Healthc. technology letters* **6**, 103–108 (2019).

47. Xia, T. *et al.* Ai-enabled microscopic blood analysis for microfluidic covid-19 hematology. In *2020 5th International Conference on Computational Intelligence and Applications (ICCIA)*, 98–102 (IEEE, 2020).

48. Wang, X., Pan, G., Hu, Z. & Ge, A. A two stage blood cell detection and classification algorithm based on improved yolov7 and efficientnetv2. *Sci. Reports* **15**, 8427 (2025).

49. Wang, N., Xi, L., Sun, Y., Chen, C. & Li, Z. The efficientnet-b3 with adaptive loss for white blood cell classification. In *2025 International Conference on Mechatronics, Robotics, and Artificial Intelligence (MRAI)*, 408–411, DOI: 10.1109/MRAI65197.2025.11135624 (2025).

50. Ferdousi, J., Lincoln, S. I., Alom, M. K. & Foysal, M. A deep learning approach for white blood cells image generation and classification using srgan and vgg19. *Telematics Informatics Reports* **16**, 100163 (2024).

51. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

52. Katar, O. & Yildirim, O. An explainable vision transformer model based white blood cells classification and localization. *Diagnostics* **13**, 2459 (2023).

53. Fu, L., Chen, J., Zhang, Y., Huang, X. & Sun, L. Cnn and transformer-based deep learning models for automated white blood cell detection. *Image Vis. Comput.* 105631 (2025).

54. Matek, C., Schwarz, S., Spiekermann, K. & Marr, C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat. Mach. Intell.* **1**, 538–544 (2019).

55. Manescu, P. *et al.* Detection of acute promyelocytic leukemia in peripheral blood and bone marrow with annotation-free deep learning. *Sci. Reports* **13**, 2562 (2023).

56. Park, S. *et al.* A large multi-focus dataset for white blood cell classification. *Sci. Data* **11**, 1106, DOI: 10.1038/s41597-024-03938-1 (2024).

57. Sarkar, R. Low distortion delaunay embedding of trees in hyperbolic plane. In *International symposium on graph drawing*, 355–366 (Springer, 2011).

58. Nickel, M. & Kiela, D. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, vol. 30 (2017).

59. Ganea, O.-E., Bécigneul, G. & Hofmann, T. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, vol. 31 (2018).

60. Peng, W., Varanka, T., Mostafa, A., Shi, H. & Zhao, G. Hyperbolic deep neural networks: A survey. *CoRR* **abs/2101.04562** (2021). 2101.04562.

61. Mettes, P., Ghadimi Atigh, M., Keller-Ressel, M., Gu, J. & Yeung, S. Hyperbolic deep learning in computer vision: A survey. *Int. J. Comput. Vis.* (2024). Advance online publication; also available as arXiv:2305.06611.

62. Gulcehre, C. *et al.* Hyperbolic attention networks. In *International Conference on Learning Representations (ICLR)* (2019).

63. Khrulkov, V., Mirvakhabova, L., Ustinova, E., Oseledets, I. & Lempitsky, V. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).

64. Gao, Z., Wu, Y., Jia, Y. & Harandi, M. Curvature generation in curved spaces for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2021).

65. Guo, Y., Wang, X., Chen, Y. & Yu, S. X. Clipped hyperbolic classifiers are super-hyperbolic classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11–20 (2022).

66. Skliar, A. & Weiler, M. Hyperbolic convolutional neural networks. *CoRR* **abs/2308.15639** (2023). 2308.15639.

67. Bdeir, A., Schwethelm, K. & Landwehr, N. Fully hyperbolic convolutional neural networks for computer vision. In *Proceedings of the International Conference on Learning Representations (ICLR)* (2024).

68. Ayubcha, C. *et al.* Improved generalizability in medical computer vision: Hyperbolic deep learning in multi-modality neuroimaging. *J. Imaging* **10**, 319, DOI: 10.3390/jimaging10120319 (2024).

69. Matek, C., Schwarz, S., Marr, C. & Spiekermann, K. A single-cell morphological dataset of leukocytes from aml patients and non-malignant controls [data set]. the cancer imaging archive. *Frederick, MD, USA* DOI: 10.7937/tcia.2019.36f5o9ld (2019).

70. Kouzehkanan, Z. M. *et al.* A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm. *Sci. reports* **12**, 1123, DOI: 10.1038/s41598-021-04426-x (2022).

71. Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. & Manag.* **45**, 427–437, DOI: 10.1016/j.ipm.2009.03.002 (2009).

72. Powers, D. M. W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2**, 37–63 (2011).

Funding

This work was supported by the Center for Artificial Intelligence and Quantum Computing in System Brain Research (CLARA) (101136607-02), the project "Research Platform for Digital Transformation and Society 5.0" (CZ.02.01.01/00/23_021/0012599), the Internal Grant Agency of VSB-TUO, Processing and Advanced Analysis of Biomedical Data XI of the Czech Republic (SP2026/008).

Author contributions

L.J. contributed to investigation, methodology, and conceptualization, and wrote the original draft. M.V. contributed to conceptualization, investigation, and writing, and performed data curation and visualization. P.G. contributed to project administration, funding acquisition, supervision, and writing (review and editing). All authors reviewed the manuscript.

Data availability

The curated 14-class white blood cell dataset, together with the source code, configuration files, and scripts required to reproduce the main analyses, is publicly available at Zenodo: https://doi.org/10.5281/zenodo.19126698. The source code is additionally available in the GitHub repository: https://github.com/lowoncuties/HemaHypNet.

The curated dataset is a derived work based on publicly available datasets. Users must comply with the original licenses and terms of use of the source datasets: AML-Cytomorphology_LMU, available from The Cancer Imaging Archive (https://www.cancerimagingarchive.net/collection/aml-cytomorphology_lmu/); Raabin-WBC, originally provided through the Raabin data availability link reported in Kouzehkanan et al., *Scientific Reports* **12**, 1123 (2022), DOI: https://doi.org/10.1038/s41598-021-04426-x; and A large multi-focus dataset for WBC classification, available from the figshare repository (https://springernature.figshare.com/articles/dataset/A_large_multi-focus_dataset_for_white_blood_cell_classification/26781052?backTo=%2Fcollections%2FA_Large_Multi-Focus_Dataset_for_White_Blood_Cell_Classification%2F6844605&file=48650791).

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- HyperbolicSRsuppl.pdf