Advancing White Blood Cell Detection: A Multi-Domain Dataset for Morphological Analysis and Addressing Sparse Annotation Challenges

Abdul Rehman, Talha Meraj, Aiman Mahmood Minhas, Ayisha Imran, Mohsen Ali, Waqas Sultani

***Abstract*— Realistic leukemia analysis on microscopic slides requires a system that localizes, classifies, and performs WBC morphological analysis. While deep learning has advanced medical imaging, leukemia analysis is hindered by the lack of a large, diverse dataset. Our contribution here is twofold. First, we present a large-scale Leukemia dataset collected through Peripheral Blood Films (PBF) from many patients, through multiple microscopes, multi-cameras, and multi-magnification. To enhance explainability and medical expert acceptance, each leukemia cell is annotated with 7 morphological attributes, ranging from Cell Size to Nuclear Shape. Specifically, we used two microscopes from two different cost spectrums (high-cost: HCM and low-cost: LCM) to capture the dataset at three magnifications (100x, 40x,10x) through different sensors (high-end camera for HCM, middle-level camera for LCM, and mobile phone camera for both). Our dataset has 2.4k annotated images per resolution (in both HCM and LCM), with 10.3K cells annotated for 14 WBC types and 7 morphological properties. Secondly, hematologists typically annotate only well-spread, non-overlapping cell regions within small fields of view (FoV), leaving larger FoVs unannotated. Models trained solely on these annotated areas underutilize available information. To overcome this limitation, we present a novel method for training object detectors using sparse annotations. Along with this, we present a second dataset Leukemia-Sparse-Attri dataset. Curated at 40x FoV, this dataset consists of 1546 sparsely annotated microscopic images, and 185 fully annotated images (used for testing). From explainability to overcoming domain-shift challenges, presented datasets could be used for many challenging aspects of microscopic image analysis.**

***Index Terms*— Explainable, Blood Cancer, Leukemia, Attribute, Sparse Annotation,**

I. Introduction

Leukemia has been reported to be 10th [1] most frequently diagnosed cancer and one of the leading causes of cancer-related deaths worldwide, with approximately 487,000 new cases and 305,000 deaths in 2022. This disease disproportionately affects people under 40 years of age, with children being particularly impacted. The diagnosis of Leukemia requires an expert hematologist with extensive experience, capable of analyzing peripheral blood slides employing costly medical equipment. Given the prerequisites for leukemia diagnosis and prognosis, many remote regions in underdeveloped countries experience higher mortality rates [2]. However, early detection of leukemia enables timely and accurate treatment, which can significantly improve survival rates. For leukemia diagnosis, hematologists examine the Peripheral Blood Film (PBF) of patients—a process that is often tedious, error-prone, and requires significant expertise. The limited availability of experts and costly diagnostic resources hinders early, precise, and timely detection of leukemia, particularly in resource-constrained areas. Furthermore, leukemia treatment requires repetitive blood sample preparation and detailed analysis requiring frequent access to experts. AI-driven solutions are appropriate to address these challenges and over the years many leukemia datasets have been collected for this task including [3]–[8]. However, these datasets suffer from many limitations, including the absence of localization annotations, a lack of association with morphological attributes, and a limited number of images. The limited dataset availability with cell-level annotations was previously considered as one of the key challenges [9]. One major limitation is the digitization of the slides, which requires identifying relevant areas on the PBF, and then photographing that in a required magnification. These variations in staining [10] make it crucial to develop a system that generalizes across staining variations and accurately recognizes both cell morphology and WBC type for reliable leukemia diagnosis. Furthermore, even after digitization, only a small portion of the acquired image is used by the expert to perform the required analysis and annotation. Hence, a large part of the effort and available information remains unused. To overcome these dataset challenges, we present a large-scale WBC dataset named 'LeukemiaAttri+', that consists of two subsets 'LeukemiaAttri' [1] and 'Sparse-Leukemia-Attri'.

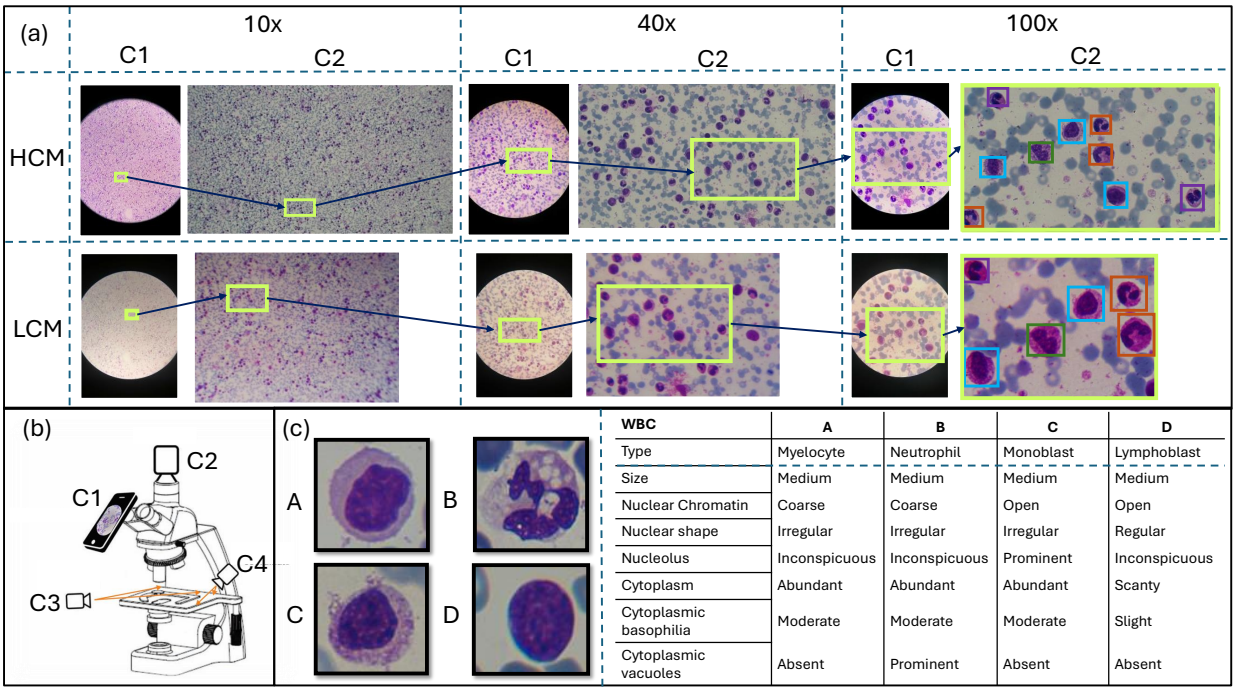
For real-world application collection, the *LeukemiaAttri* dataset was captured using two microscopes from different

Abdul Rehman, Talha Meraj, Mohsen Ali, and Waqas Sultani are with the Intelligent Machine Lab, Information Technology University of Punjab, Lahore, Pakistan (email: phdcs23002@itu.edu.pk; talha.meraj@itu.edu.pk; mohsen.ali@itu.edu.pk; waqas.sultani@itu.edu.pk).

Aiman Mahmood Minhas and Ayisha Imran are with the Department of Hematology, Chughtai Lab Lahore, Pakistan (email: dr.aiman.6258@cll.edu.pk; dr.ayisha@cll.edu.pk).

Dataset and codes are shared publicly at: https://tinyurl.com/586vaw3j.

[1]This subset was published in *'A Large-scale Multi Domain Leukemia Dataset for the White Blood Cells (WBC) Detection with Morphological Attributes for Explainability'* at MICCAI 2024 [11], and this journal paper serves as its extension.



WBC	A	B	C	D
Type	Myelocyte	Neutrophil	Monoblast	Lymphoblast
Size	Medium	Medium	Medium	Medium
Nuclear Chromatin	Coarse	Coarse	Open	Open
Nuclear shape	Irregular	Irregular	Irregular	Regular
Nucleolus	Inconspicuous	Inconspicuous	Prominent	Inconspicuous
Cytoplasm	Abundant	Abundant	Abundant	Scanty
Cytoplasmic basophilia	Moderate	Moderate	Moderate	Slight
Cytoplasmic vacuoles	Absent	Prominent	Absent	Absent

Fig. 1. a) Illustrates the image-capturing process using general mobile cameras (C1) high-end cameras for HCM, and middle-level cameras for LCM (C2) at multiple resolutions. Images are captured using both high and low-cost microscopes. b) Shows our microscope set up with cameras (C1, C2) to capture slide images and cameras (C3, C4) to capture the stage-scales. c) Shows different WBC types with their morphological attributes.

price ranges—one high-cost (HCM) and one low-cost (LCM). The data was obtained at three magnifications (100x, 40x, 10x) using various sensors: a high-end camera for HCM, a mid-range camera for LCM, and a mobile phone camera for both microscopes. The high-end camera is 47 times more expensive than the middle-range camera and HCM is 17 times more expensive than LCM. Although LCM is significantly cheaper than HCM, it sacrifices visual details and offers a more restricted FoV. In this collection, using HCM at high resolution (100x), experienced hematologists annotated 10.3k WBC of 14 types, having 55k morphological labels from 2.4k images of several PBS leukemia patients. Later on, these annotations are transferred to other two magnifications (40x and 10x) of HCM, and three magnifications of LCM (100x, 40x and 10x), and on each camera captured images. The procedure of dataset collection is illustrated in Fig.1. In this dataset, morphological attributes were initially annotated at a small FoV using a 100x magnification by hematologists to ensure error-free annotations. Note that, the smaller the FoV, the smaller portion of a slide visible to the hematologist, hence more patches are needed to digitize and review the full slide. These annotations were then transferred to subsets with 40x and 10x resolution. Since annotations were mapped from 100x (small FoV) images to corresponding regions in the 40x (larger FoV) resolution subsets, they only cover about 20 percent of the area in the 40x images due to differences in FoV caused by the varying resolutions. While collecting *LeukemiaAttri*, we observed that medical experts prefer to analyze the PBF at 40x magnification due to the larger FoV and that the *LeukemiaAttri* dataset utilized only a portion of the digitized imagery. Since the region visible at 100x magnification is only 20% of the area visible at 40x magnification, only this smaller area was included in the *LeukemiaAttri* dataset. As a result, the subset of images captured at 40x resolution remains sparsely annotated, as illustrated in Figure 2.

To properly utilize the scarce resource, we curated a dataset called *Sparse-Leukemia-Attri* to train the Sprase Object Detector. For that, we have expanded the training dataset to encompass the complete 40x FoV, ensuring that the model is exposed to a broader spectrum of morphological features. For the training set only 20% of the image is annotated, however, for evaluation on the testing set, the full image is annotated. Sparse-Leukemia-Attri is collected only on one microscope and one sensor. We combine these two sub-datasets under one label, the Leukemia+ dataset. To show the efficacy of our collected dataset, YOLOv5 is modified by integrating a lightweight head, named AttriHead, specifically designed for predicting WBC attributes. This model is labeled as AttriDet. Building on AttriDet, we propose a method for the detection of leukemia cells and their attributes through sparse annotation. This model is named Sparse Leukemia-Attribute Detector (SLA-Det). SLA-Det incorporates region-guided masking and pseudo-label selection based on multiple criteria, ensuring robust model performance. We believe these extensions make our dataset and framework more aligned with clinical practices and enhance the reliability of automated leukemia detection. In summary 1) we present a large-scale, multi-domain LeukemiaAttri dataset, utilizing two different microscopes (HCM and LCM), two cameras (C1 and C2), and three resolutions (10x, 40x, 100x). In addition, we propose *Sparse-Leukemia-Attri* by including full FoV of 40x for both training and testing, where testing data is fully annotated and

training data is sparsely annotated 2) We propose two methods to detect WBC along with their morphological attributes. The first method, AttriDet, detects WBC and morphological attributes in a fully supervised way, while the second method, SLA-Det detects the same while being trained on sparsely labeled data. 3) To effectively integrate information from labeled and unlabeled regions, during SLA-Det training, we propose several problem-specific criteria to improve better pseudo-label selection. These involve constraints on size, objectness score, and entropy of the class labels. 4) SLA-Det is around four times more efficient than AttriDet [11], while keeping the same detection accuracy, making it more useful for diagnostic purposes.

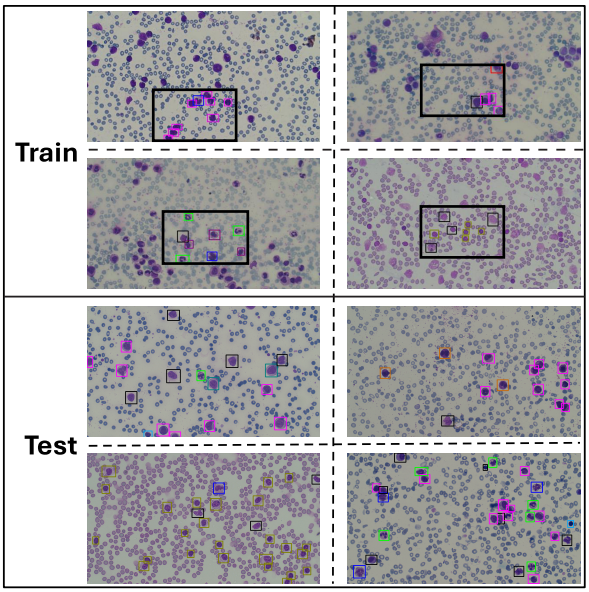


Fig. 2. Sparse-Leukemia-Attri: The training data has sparse annotations, while the test data is fully annotated. Both are at 40x magnification.

II. Related Work

Our dataset includes densely labeled patches and sparsely labeled large images, allowing for both fully supervised Object Detection (OD) and Sparsely Annotated Object Detection (SAOD). Below is a brief review of these methods, highlighting their relevance to WBC detection.

Supervised WBC detection: Deep learning (DL) techniques have been extensively applied to WBC classification [12]–[14]. Huang *et al.* [15] implemented a two-step approach for their end-to-end solution: first, detecting the WBCs, and then performing classification after cropping the detected cells. To classify single-cell images, Saidani *et al.* [16] optimized a CNN model and employed multiple pre-processing techniques, such as color conversion and image histogram equalization. Similarly, authors in [17] performed the classification of two types of WBCs from the ALL-IDB dataset [17]. G Liang et al. [18] developed a combined model of convolutional neural networks (CNN) and recurrent neural networks (RNN) to classify four types of WBCs using the BCCD dataset [19].

To detect WBCs in whole microscopy images, the authors in [20] gave a lighter deep-learning CNN model and adapted the YOLOX-nano [21] baseline to identify the WBC types. Their method enhances the overall model architecture by improving Intersection over Union (IoU) loss. A blood film and bone marrow aspirates-based multiple instance learning for leukocyte identification approaches have been proposed to classify different types of Leukemia [22]. For image preprocessing, they mainly applied thresholding methods. The diagnostic-level full images of PBF are used to train and test the model and the object-level dataset [3] is used to validate the model.

Sparsely annotated object detection: To tackle the challenge of object detection in sparsely annotated datasets, authors in [24] proposed a method that integrates supervised and self-supervised learning into their framework. The approach involves extracting features from an image and its multiple augmented versions, which are then processed through a shared region proposal network. A positive pseudo-mining technique is introduced to effectively filter out labeled, unlabeled, and foreground regions. This results in separate losses for labeled regions—supervised loss and self-supervised loss—ensuring consistent feature extraction across both original and augmented views. This combined method enhances the model's generalization ability and improves object detection performance. Similarly, Z. Wu *et al.* [25] used a soft sampling technique to mitigate the impact of missing annotations. Their approach applies the Gompertz function to adjust gradient weights based on the overlap between positive instances and regions of interest. In another research work [26], part-aware learning is performed by training a model twice. This approach leverages spatial relationships between part and subject categories to selectively ignore classification loss for part categories when they are likely included within subject categories. It employed the training of the model twice and used pseudo labels generated from the first model to guide the training of the second model. These methods aim to enhance model training and performance in the presence of sparse annotations. To generate labels for sparsely annotated datasets, the co-mining approach is applied by authors of [27]. It involves constructing a Siamese detection network with two branches, each generating pseudo-label sets using a co-generation module. These pseudo labels based and complete labels sets based new set is prepared. It is subsequently used as guided signals to the model, leveraging augmentation techniques to generate diverse pseudo-label sets. Finally, a co-generation module converts predictions from one branch into pseudo-labels for the other, facilitating complementary supervision between the network branches. Note that previous benchmarks, such as [24], [27], implemented sparse object detection algorithms on self-created datasets. In contrast, our dataset consists of small patches annotated with detailed morphological features and patch localization. In addition, we introduce a multi-headed WBC detection method tailored for such data.

III. Dataset

Leukemia comprises four main types: Chronic Myeloid Leukemia (CML), Chronic Lymphocytic Leukemia (CLL), Acute Lymphocytic Leukemia (ALL), and Acute Myeloid

Leukemia (AML) [1]. These types fall into two primary lineages, myeloid and lymphoid, with abnormalities leading to atypical CLL lymphocytes, ALL lymphoblasts, CML myeloblasts, and elevated AML myeloblasts.

A. Limitations of existing WBC datasets

The details of publicly available datasets on WBC and leukemia are shown in Table I, which highlights various limitations such as: 1) The ALL-IDB dataset [3] contains only 510 white blood cells (WBCs) from patients with Acute Lymphocytic Leukemia (ALL) and includes only cell centroid information. 2) In version 2 of the ALL IDB [4], each image contains only a single WBC, with the cell's status labeled accordingly. 3) The LISC [5] dataset contains a smaller number of WBCs with five classes such as basophil, eosinophil, lymphocyte, monocyte, and neutrophil. 4) The entire dataset of Munich AML's [6] contains single-cell images that come under the umbrella of AML. 5) Only Raabin [7], out of all the aforementioned existing datasets, has complete bounding box-level annotation with single-cell images, but the entire Raabin-WBC dataset is made up of normal individuals. Furthermore, the Raabin-WBC collection contains images from two separate cameras and two different microscopes, without maintaining any correspondence across microscopes. 6) The HRLS [8] dataset is captured using a single microscope and covers only two types of Leukemia and nine types of WBCs. 7) WBCAtt dataset [23] is the improved version of [28], where annotations of attributes are added for each WBC. The dataset [28] lacks the abnormal WBCs and is captured using a single microscope only.

B. LeukemiaAttri dataset

To address dataset limitations and establish a benchmark for deep learning models, we collected the 'LeukemiaAttri' dataset. Healthcare professionals define the key WBC types and their morphological attributes for a comprehensive understanding of leukemia prognosis. In this dataset collection, the PBFs are collected from the diagnostic lab and images are captured from the monolayer area. To capture images, we utilize two distinct microscopes – the high-cost (Olympus CX23) and the low-cost (XSZ-107BN) – in conjunction with two separate cameras, namely the HD1500T (HCM), ZZCAT 5MP (LCM) and the Honor 9x Lite mobile camera (HCM, LCM). It is quite challenging to locate the same patch on the PBF when employing different microscopes and resolutions [29]. To address this inherent challenge, we initiated

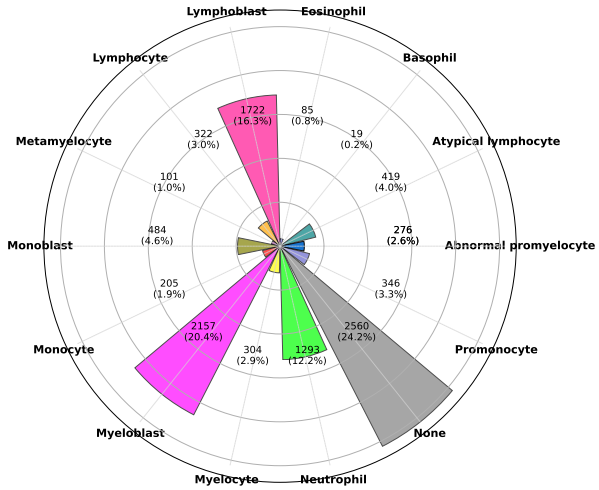


Fig. 3. WBC type distribution in the LeukemiaAttri dataset.

the capturing process by setting the FoV at 10x, and 40x with an approximate 20% overlap, maintaining a fixed x-axis stage scale. At 100x magnification, we captured the FoV containing WBCs without any overlap, ensuring the distinct representation of individual WBCs. This process is repeated both for HCM and LCM. The *LeukemiaAttri* contains 14 types of WBC cells along with the 'None' category containing artifacts, bear cells, and cells that were difficult to recognize by hematologists. Overall, the LeukemiaAttri dataset consists of 28.9K (2.4K $\times$ 2 $\times$ 3 $\times$ 2) images captured using low-cost and high-cost microscopes at three different resolutions (10x, 40x, 100x) and different cameras. This way, we have 12 subsets of images.

Morphological attributes: To enhance prognostic assistance, hematologists identified the 14 types of WBCs. Morphology details of some cells are shown in Figure 1(c). To annotate the WBC type and morphology attributes, hematologists reviewed all subsets of the captured images. They then selected a subset containing the best-quality images with detailed structural information for annotation. This subset was collected at 100x using an HD1500T camera on HCM (HCM_100x_C2). For quality control, two hematologists annotate cells with the consultation. The detail of some types of WBC with the morphological information is shown in Fig 1 (c), where A) monoblasts, B) monocytes, and C) myelocytes cells exhibit mostly similar morphological characteristics as they originate from the myeloid lineage. Nevertheless, differences arise, particularly in the presence of cytoplasmic vacuolation. However,

TABLE I
COMPARISON OF THE PROPOSED DATASET WITH EXISTING WBCS DATASETS

Dataset	Type	Across Micro.	Multi. Cells in image	BBX	Multi Resolution	No. of WBC's	WBC Classes	Morphology
IDB [3]	ALL	✘	✔	✔	✘	510 (LB)	2	✘
IDB2 [4]	ALL	✘	✘	✘	✘	260	2	✘
LISC [5]	Multi.	✘	✘	✘	✘	250	6	✘
Munich [6]	AML	✘	✘	✘	✘	18.3K	15	✘
Raabin [7]	Normal	✘	✔	✔	✘	17.9K	5	✘
HRLS [8]	Multi.	✘	✔	✘	✘	16K	9	✘
WBCAtt [23]	Normal	✘	✘	✘	✘	10.2K	5	✔
Ours [2]	Multi.	✔	✔	✔	✔	**88.2K**	14	✔

D) lymphoblasts belong to a lymphoid lineage that shows morphological dissimilarities in both lineages. After obtaining detailed WBC and their attributes annotations from hematologists for HCM at 100x, we transferred the annotations to different resolutions and across the microscope automatically using homography [30], [31]. Transferred annotations were verified manually and re-annotatation was done for the missing localization. The detailed count of the source subset of WBC types and their corresponding attributes are shown in Fig. 3.

C. Sparse LeukemiaAttri dataset

Based on our experiments (see Table II), the best detection results were achieved with the high-cost microscope and C2 (HD1500T). Additionally, despite the hematologists' preference for quickly analyzing a complete 40x FoV, it is noteworthy that the entire 40x FoV is not included in the LeaukeiaAttri dataset. To address the limitations of the LeukemiaAttri dataset, we extend the HCM_40X_C2 subset to include a complete FoV. This is named as *Sparse-Leukemia-Attri*. To save the hematologist time and effort, we obtain a whole FoV annotation of the testing set only and for the training set, we rely on approaches that learn the model from the sparsely labeled datasets (see Figure 2). The Sparse LeukemiaAttri dataset is then split up into train (1546 images) and test sets (185 images) whereas the hematologist annotates the test set at 100x (for better annotation quality) resolution and these annotations are transferred to a 40x test set. The ratio of the labeled patch to whole FoV WBC both in size and number of WBCs is 1:5. Collectively, with a new extension, we have 13 subsets now as shown in Table 1 of Supplementary. Our *Sparse-Leukemia-Attri* dataset is a specialized case of sparsely supervised and semi-supervised object detection datasets. Unlike traditional sparsely supervised datasets where annotations are randomly missed, each of our images contains a fully labeled patch. Similarly, while semi-supervised approaches typically involve fully labeled and fully unlabeled sets, our dataset consists of partially annotated patches within each image, rather than entirely labeled or unlabeled images.

IV. Methodology

A. Preliminaries

Let $\mathcal{F}(I_i) = \{\hat{\mathbf{z}}_{i,k}\}_{k=1}^{n_i}$ be the object detector that outputs n_i predictions. Each prediction $\hat{\mathbf{z}}_i^k = (\hat{c}_{i,k}, \hat{o}_{i,k}, \hat{\mathbf{p}}_{i,k}, \hat{\mathbf{b}}_{i,k})$ consists of objectness score $\hat{o}_{i,k}$, a class with maximum probability $\hat{c}_{i,k}$, predicted probability vector $\hat{\mathbf{p}}_{i,k}$, and bounding-box $\hat{\mathbf{b}}_{i,k} \in \mathbb{R}^4$. The confidence score, $\hat{p}_{i,k}$, for k^{th} prediction is computed as: $max(\hat{\mathbf{p}}_{i,k}) \times \hat{o_{i,k}}$, where max returns the maximum value from the input vector. For readability, we assume that $\hat{\mathbf{b}}_{i,k} = \{x_k, y_k, w_k, h_k\}$ and $I_i \in \mathbb{R}^{H\times W}$ is the input image and $\{\mathbf{z}_{i,k}\}_{k=1}^{n_i}$ is the annotation of the input image. Let $\mathcal{D}_{\mathcal{A}} = \{(I_i, \{\mathbf{z}_{i,k}\}_{k=1}^{n_i})\}_{i=1}^{N_D}$ be the fully annotated dataset. Each annotation $\mathbf{z}_i^k = (\mathbf{b}_{i,k}, \mathbf{c}_{i,k}, \mathbf{m}_{i,k})$ consists of a bounding box $\mathbf{b}_{i,k}$, a corresponding class label $\mathbf{c}_{i,k} \in \{1, \ldots, C\}$, and $\mathbf{m}_{i,k}$, which is the set of morphological attributes associated with the kth object in the ith image.

Our methodology is compatible with any object detector, however, we chose YOLOv5 [32] for our approach, since our experiments indicated (Table II) it to have higher mAP in comparison with other recent SOTA object detectors including FCOS [33], Spare-Faster-RCNN [34] and transformer-based DINO [35].

B. AttriDet: WBC detection with attributes

Our objective is to augment existing object detector with morphological attribute detection module. Let $\mathcal{B}$ be the feature extractor part of the object detector $\mathcal{F}$. In order to capture the structure and low-level features that are necessary for morphological property recognition, we extract N feature maps $\{S_j\}_{j=1}^{N}$ from the initial layers. For $I_i \in \mathbb{R}^{H\times W}$ as an input image, assume j^{th} feature map as $S_j \in \mathbb{R}^{H_j\times W_j\times C_j}$. In order to extract features of kth object using the ground-truth bounding-box $\mathbf{b}_{i,k} = \{x_k, y_k, w_k, h_k\}$ during training and predicted bounding box $\hat{\mathbf{b}}_{i,k}$ during the testing, bounding box has to be normalized per the size of feature map S_j and image size.

The normalized bounding box for $\mathbf{b}_{i,k}$ is given by:

$$\mathbf{b}_{\text{norm}}^{S_j,k} = \left(\frac{x_k}{W_i}W_j, \frac{y_k}{H_i}H_j, \frac{w_k}{W_i}W_j, \frac{h_k}{H_i}H_j\right), \tag{1}$$

The normalized bounding box $\mathbf{b}_{\text{norm}}$ is used to extract features from S_j followed by ROIAlign [36] to equalize the dimensions and properly align the extracted feature. Let $\mathcal{ROI}$ denote the ROIAlign operation.

$$\mathcal{F}_{S_j,k} = \mathcal{ROI}(S_j, \mathbf{b}_{\text{norm}}^{S_j}) \quad \text{for } j \in \{1, 2\}. \tag{2}$$

In our case, dimensions of $\mathcal{F}_{S_1,k}$ and $\mathcal{F}_{S_2,k}$ are 24×30×128 and 24×30×256, respectively. Finally, multi-scale features are fused $\mathcal{C}_{\mathcal{F}}$=$\mathcal{F}_{S_{1,k}} \oplus \mathcal{F}_{S_{2,k}}$ and then fed into the AttriHead to effectively learn the relevant WBC attributes. In AttriHead, three convolutional layers are followed by max pooling layers. The output is then flattened into a d-dimensional vector and passed through a fully connected (FC) network with three layers configured as (1024, 256, 6) units. The training of AttriHead is done using multilabel binary class loss via asymmetric loss [37]. Once trained, the AttriHead predicts the six-dimensional vectors $\hat{\mathcal{M}}^{1\times 6}$, where each dimension represents the state of the attribute. For the morphology detail, please refer to the section III. In addition to the AttriHead, we also train a YOLOv5 head for WBC detection, utilizing the following loss functions with the standard settings of the detector:

$$\mathcal{L}_{total} = \sum_{l=1}^{N} \left(\mathcal{L}_{obj}^{l} + \mathcal{L}_{cls}^{l} + \mathcal{L}_{bbox}^{l} + \mathcal{L}_{mor}\right) \tag{3}$$

where l denotes the output prediction layers, $\mathcal{L}_{total}$ represents the final loss of the object detector, which is composed of several components: $\mathcal{L}_{\text{cls}}$ represents the object classification loss, $\mathcal{L}_{\text{bbox}}$ denotes the bounding box regression loss, $\mathcal{L}_{\text{obj}}$ is the loss associated with objectness and $\mathcal{L}_{mor}$ for the attribute classification. Overall the proposed AttriDet not only detects WBCs but also provides the detailed morphology in terms of attributes which is useful for explainbility.

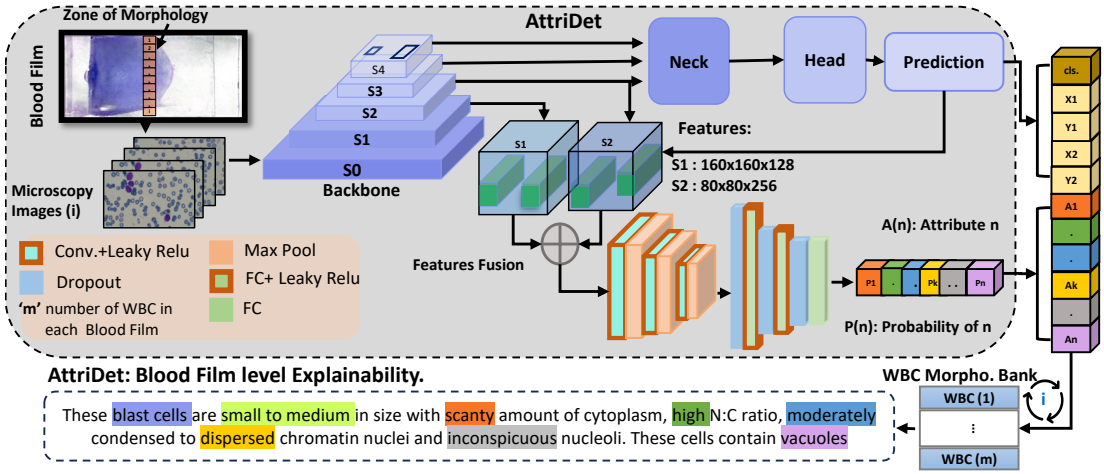


Fig. 4. AttriDet: WBC and their corresponding attribute predictions framework

C. SLA-Det: WBC detection with attributes using Sparse Annotation

Since many object detectors (e.g. Yolov5) use the IoU of the predicted bounding box against ground truth to identify the true positives and calculate the object loss, training with sparse annotations negatively impacts performance because the ground truth data restricts the model's ability to learn effective features, causing it to struggle with differentiating between background and objects of interest. This can increase false negatives, as the model may overlook unannotated objects. To overcome these limitations sparse-annotation-based object detectors, such as SparseDet [24] and Co-mining [27], have been proposed. Usually, these methods utilize a few bounding-box annotations of the objects (with many objects present but not annotated), without much information about the background. However, the scenario on our hand is slightly more informative and requires re-designing of the solution. In our case, we have annotated regions, providing us with knowledge of where objects are present and where they are absent within those areas. Outside that region, we have no information about objects and backgrounds. For this, we propose a new sparse-annotation-based object detector of Leukemia, called Sparse Leukemia Attribute Detector (SLA-Det). Let $\mathcal{D}_{\mathcal{S}} = \{(I_i, \{\mathbf{z}_{i,k}\}_{k=1}^{n_i}, P_i)\}_{i=1}^{N_D}$ be that sparsely annotated dataset. Similar to $\mathcal{D}_{\mathcal{A}}$, each annotation $z_i^k = (\mathbf{b}_{i,k}, \mathbf{c}_{i,k}, \mathbf{m}_{i,k})$ consists of bounding box $\mathbf{b}_{i,k}$, corresponding class-label $\mathbf{c}_{i,k} \in \{1, \ldots, C\}$ and $\mathbf{m}_{i,k}$ is the set of morphological attributes associated with the kth object in the ith image. With each image I_i we have associated patch P_i indicating fully annotated region. Note that this region is about 20% of the image (as shown in Figure 2). Let $\mathcal{F}(I_i) = \{\hat{\mathbf{z}}_{i,k}\}_{k=1}^{n_i}$ denote the object detector that generates n_i predictions for the image I_i. The predicted output could be separated into non-intersecting sets $\hat{\phi}_i$ and $\hat{\varphi}_i$. Where $\hat{\phi}_i = \{$set of all $\hat{\mathbf{z}}_{i,k}$ that do not lie inside $P_i\}$ and $\hat{\varphi}_i = \{$set of all $\hat{\mathbf{z}}_{i,k}$ that lie inside $P_i\}$.

1) Training on Annotated Region : For the $\hat{\varphi}_i$ that lies inside the annotated patch P_i, we can leverage the ground truth to identify correct detections and penalize any detections that occur over the background. Let M_i be the mask generated using P_i, such that it's zero outside the P_i and 1 inside.

Handling Multiple Prediction Layers: Several object detectors ([38]–[40]) generate predictions at multiple levels / layers $\{l^v\}_{v=1}^{N}$, to detect objects at multiple scales. In order to penalize the objectness score assigned to the background region, the location of the annotated patch has to be adjusted to these layers. We scale P_i for l^v by applying normalization (Equation 1) creating $P_{i,\text{norm}}^v$. For each level, we separate the predictions, such that $\forall_v \hat{\varphi}_i^v = \{$set of all $\hat{\mathbf{z}}_{i,k}$that lie inside $P_i^v\}$.

Labeled region loss: To fully utilize the information in the annotated patch, the objectness loss has to be backpropagated. Since objectness is different at different layers (as each layer predicts objects of different sizes), we compute the masked objectness loss for each layer using the predication of $\hat{\varphi}_i^v$ that lies within the annotated patch. This ensures that the loss calculation focuses only on relevant regions, allowing the model to better learn from the annotated objects while ignoring irrelevant background information. Final loss for annotate region is Eq. (4):

$$\mathcal{L}_{LR} = \sum_{v=1}^{N} \left(\mathcal{L}_{obj}(\hat{\varphi}_i^v) + \mathcal{L}_{cls}(\hat{\varphi}_i^v) + \mathcal{L}_{bbox}(\hat{\varphi}_i^v) \right), \quad (4)$$

2) Training on non-Annotated Region: To address the lack of annotations in unlabeled regions, we rely on a self-training approach by selecting pseudo-labels from the predictions in the non-annotated region $\hat{\phi}_i$.

Pseudo-Label filtering: Since predictions are noisy and might be incorrect, we design filtering criteria. First, we filter out predictions, $\mathbf{z}_{i,k}^*$, that have predicted confidence score, $p_{i,k}^*$ below θ_0. This effectively means that only predictions from unlabeled regions with a high enough confidence, considering both the class probability and object score, are retained for further processing. Secondly, we remove all object predictions with an area smaller than the smallest object (Area$(\hat{b}_{i,k}) >$ Area$_{\text{min}}$) in the ground-truth in the training set. Thirdly, in the remaining set, for any prediction with confidence score $\hat{p}_{i,k}^*$ above θ_1, we compute entropy, allowing us to focus on the most distinguished class probabilities [41].

$$\mathcal{E}_{i,k} = -\sum_{c' \in \mathcal{C}} \hat{\mathbf{p}}_{c'} \log_2 \hat{\mathbf{p}}_{c'}, \quad (5)$$

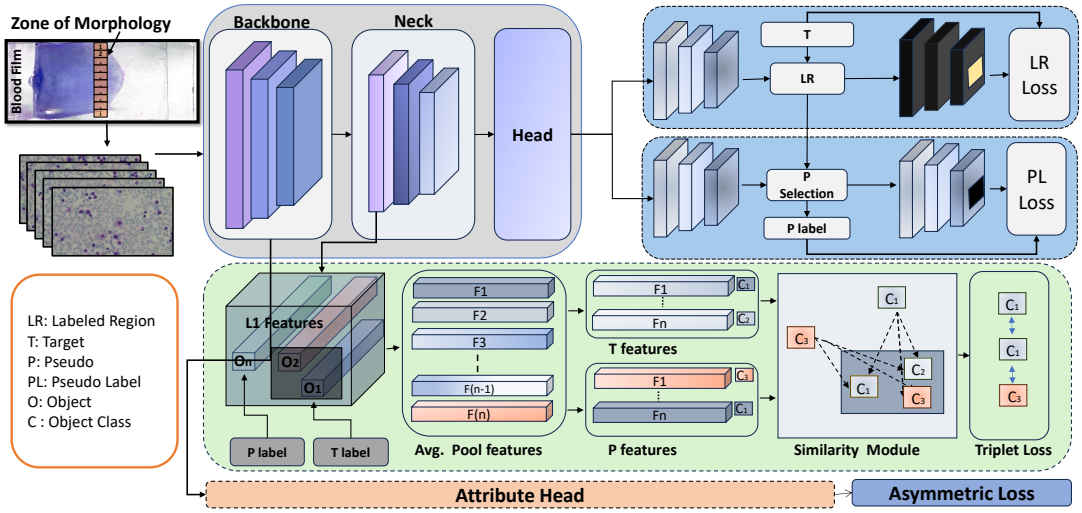


Fig. 5. SLA-Det: WBC and their corresponding attribute predictions framework for the sparsely annotated microscopy images

where $\mathcal{C}$ is the set of all class probabilities of $c^*_{i,k}$, and $\hat{\mathbf{p}}_{c'}$ is the probability of the outcome c'. In summary,

$$D_i^{pseudo} = \{\hat{\mathbf{z}}_{i,k} \in \hat{\phi}_i \mid \text{Area}(\hat{b}_{i,k}) > \text{Area}_{\text{min}} \;\&\; \hat{c}_{i,k} > \theta_1 \;\&\; \mathcal{E}_{i,k} < \theta_2\}, \tag{6}$$

where θ_2 denotes the entropy threshold. The above-selected pseudo-labels could not be used to reject the background as they only tell where the object is but not where an object could not be. Therefore, we need to mask out the object loss for the regions outside the pseudo-label. For $\hat{\mathbf{z}}_{i,k} \in D_i^{pseudo}$, we define the pseudo-label $\mathbf{z}_i^{*k} = (\hat{c}_{i,k}, \hat{\mathbf{b}}_{i,k}, M_{i,k})$, where $M_{i,k}$ is mask of image size with only object location set to 1, rest is zero. Let $\phi_i^* = \{\mathbf{z}_i^{*k} | \hat{\mathbf{z}}_{i,k} \in D_i^{pseudo}\}$. For each level, the elements in the set are normalized per the size of the level and result in sets ϕ_i^{*v}.

$$\mathcal{L}_{PL} = \sum_{v=1}^{N} \left(\mathcal{L}_{\text{obj}}(\hat{\phi}_i^v) + \mathcal{L}_{\text{cls}}(\hat{\phi}_i^v) + \mathcal{L}_{\text{bbox}}(\hat{\phi}_i^v) \right), \tag{7}$$

where the $\mathcal{L}_{PL}$ represent the loss from the unlabeled region. **Similarity Based Training:** We hypothesize that if the classes of two objects (in our case, the object class from the labeled region versus the unlabeled region) are the same, then their deep features should be very similar. Conversely, if the classes are not the same, their features should not be similar. Therefore, the minimum similarity between objects of the same class must be greater than the maximum similarity between objects of different classes. Based on this hypothesis, we design a novel strategy to use $\hat{\mathbf{z}}_{i,k} \in \hat{\phi}_i$ which have confidence score greater than θ_0 indicating they might not be background but less than θ_1 (indicating we are not sure of the class label). Let $\hat{\mathbf{z}}^s_{i,k}$ be such predictions. For $\hat{\mathbf{z}}^s_{i,k}$, we extract the features, $\hat{f}_{i,k}$, from the initial layers of neck. Similarly, for each jth ground-truth bounding box in annotated patch (P_i), we extract $f_{i,j}$. To enhance adaptability, the triplet difference is determined by evaluating the feature maps of each filtered label against all ground truth feature maps within the batch:

$$\text{Min_Sim}(\hat{f}_{i,k}) = \min_{p \in f_i} \left(\frac{p \cdot \hat{f}_{i,k}}{\|p\| \|\hat{f}_{i,k}\|} \right), \text{for all } \hat{c}^s_{i,k} = \hat{c}_{i,j} \tag{8}$$

$$\text{Max_Sim}(\hat{f}_{i,k}) = \max_{p \in f_i} \left(\frac{p \cdot \hat{f}_{i,k}}{\|p\| \|\hat{f}_{i,k}\|} \right), \text{for all } \hat{c}^s_{i,k} \neq \hat{c}_{i,j} \tag{9}$$

$$Tri_{i,k} = \max(0, \text{Min_Sim}(\hat{f}_{i,k}) - (\text{Max_Sim}(\hat{f}_{i,k}) + 0.05)). \tag{10}$$

The total triplet loss is computed as:

$$\mathcal{L}_{Tri} = \frac{1}{m} \sum_{k=1}^{m} Tri(\hat{f}_{i,k}, f_{i,j}), \tag{11}$$

where m represent the numbers of object features in $\hat{f}_{i,k}$. The final loss function for learning to detect leukemia WBCs in sparsely annotated settings is as follows:

$$\mathcal{L}_{SL} = \mathcal{L}_{LR} + \mathcal{L}_{PL} + \mathcal{L}_{Tri} \tag{12}$$

Similar to AttriDet, SLA-Det is built upon the YOLOv5 object detector (Fig. 5), however, the presented methodology for the AttriHead is adopted for the attribute prediction with the same setting and only for the region where the ground truth is available. Finally, in order to learn to predict attributes (using AttriHead) along with WBC detections, we employ the following loss.

$$\mathcal{L}_{SLA} = \mathcal{L}_{LR} + \mathcal{L}_{PL} + \mathcal{L}_{Tri} + \mathcal{L}_{mor} \tag{13}$$

V. Experiments and Results

A. Experiments on LeukemiaAttri

Implementation details: For AttriHead, dropout regularization (20%) is applied between FC layers, with Leaky ReLU activation used for both convolutional and FC layers. AttriDet is trained on the LeukemiaAttri dataset with specified augmentations and loss functions. We set the IoU threshold at 0.20 and a mosaic augmentation probability of 1.0. The optimizer

TABLE II
OBJECT DETECTION BASELINES RESULTS ON *LeukemiaAttri* DATASET

Method	Subset	$mAP@_{50-95}$	$mAP@_{50}$
Sparse R-CNN [34]	HCM_100x_C1	15.8	29.6
	HCCM_40x_C1	13.7	27.0
	HCM_100x_C2	21.3	36.7
	HCM_40x_C2	18.4	32.7
FCOS [33]	HCM_100x_C1	16.4	31.8
	HCM_40x_C1	12.6	24.8
	HCM_100x_C2	22.5	40.6
	HCM_40x_C2	17.0	32.7
DINO [35]	HCM_100x_C1	17.0	33.8
	HCM_40x_C1	22.1	36.1
	HCM_100x_C2	25.4	43.7
	HCM_40x_C2	20.9	36.9
YOLOv5x [32]	HCM_100x_C1	20.9	**38.8**
	HCM_40x_C1	19.7	**36.1**
	HCM_100x_C2	26.3	**44.2**
	HCM_40x_C2	20.1	**37.3**

TABLE III
TESTING SET RESULTS OF ATTRIDET AND SOTA METHODS ON HCM_100X_C2 SUBSET: WBC TYPE, ATTRIBUTES (NC: NUCLEAR CHROMATIN, NS: NUCLEAR SHAPE, N: NUCLEUS, C: CYTOPLASM CB: CYTOPLASMIC BASOPHILIA CV: CYTOPLASMIC VACUOLES)

Method	F1						WBC
	NC	NS	N	C	CB	CV	$mAP@_{50-95}$
CBM [42]	21.9	96.2	41.8	77.2	70.2	3.33	27.6
AttriDet [11]	**73.9**	**95.9**	**54.3**	**89.7**	**83.6**	**29.1**	**28.2**

TABLE IV
DOMAIN ADAPTATION RESULTS FOR OBJECT DETECTION ON THE *LeukemiaAttri* DATASET

Method	$mAP@_{50-95}$	$mAP@_{50}$
YOLOv5 [32]	11.0	25.5
DACA [43]	12.6	30.2
ConfMix [34]	12.6	33.5

settings include a weight decay of 0.0005, momentum of 0.937, and warmup biases of 0.1. Training is conducted for 100 epochs with a batch size of 4 on an NVIDIA GTX 1080 GPU, using a learning rate: 0.01 to 0.001) and a 3-epoch warmup.

WBC Detection: To evaluate WBC detection performance on *LeukemiaAttri*, we perfom experiments with several well-known object detectors including Sparse R-CNN [34], FCOS [33], DINO [35], and YOLOv5x [32]. We have performed detailed experiments on the dataset collected using low-cost microscope (LCM) and high-cost microscope (HCM) employing mobile cameras (C1) and special purpose microscopic camera (HD1500T for HCM and MZZCAT 5MP for LCM - C2) on 40x and 100x magnifications. The experimental results are shown in Table II. The results indicate that the best WBC detection accuracy (mAP$@_{50-95}$ of 26.3 and mAP$@_{50}$ of 44.2) is achieved on the data collected using HCM at 100x employing C2 camera using YOLOv5x. Similarly, when comparing on 40x, YOLOv5x outperforms (mAP$_{50\text{-}95}$ of 20.1 and mAP$_{50}$ of 37.3) other objection detection methods. A similar pattern can be observed for other subsets as well. We believe this is due to the YOLOV5 robust feature learning for different sizes of cells in the LeukemiaAttri dataset. Given that YOLOv5x demonstrated superior performance across both the HCM and LCM subsets of microscope and mobile camera data, we extend this for attribute prediction.

WBC detection with attribute prediction : Table. III demonstrate results of our proposed AttriDet for attributes prediction. In addition, we demonstrate the improved WBC detection of AttriDet (last column) as compared to standard YOLOv5. We credit the enhanced WBC results to robust feature learning using the attributes head. Findings from CBM [42] and AttriDet demonstrate superior performance in predicting WBC types and their attributes. However, detecting nucleus and cytoplasmic vacuoles remains a significant challenge.

Unsupervised Domain Adaptation based WBC detection: In the LeukemiaAttri dataset, 12 subsets are collected from different domains via HCM and LCM. These different domain subsets contain challenging domain shifts as shown in Fig. 1 (a), making our dataset a domain adaption benchmark. Note that due to the poor image quality of LCM, hematologists are often reluctant to provide annotations, as the process is tedious and prone to errors. Therefore, UDA methods could be used to train object detectors on high-quality images of precise annotation (HCM) and provide results on LCM. To provide the baselines of UDA, we have experimented with two recent methods; ConfMix [34] and DACA [43], utilizing the highest resolution (100x) subsets collected via HCM and LCM (source data: HCM_100x_C2 and target data: LCM_100x_C2). As can be seen in Table IV, YOLOv5x (source only) was trained on the HCM subset, achieving a 25.5 mAP$@_{50}$ on a comparable subset of the LCM. Nevertheless, employing UDA methods such as ConfMix and DACA led to higher mAP$@_{50}$ of 33.5 and 30.2, respectively. The low performance of state-of-the-art UDA methods highlights the complexity and substantial domain shift present in our dataset.

B. Experiments on Sparse-Leukemia-Attri

Implementation details: For the SLA-Det implementation, we set the values as follows: θ_0 : 0.70, θ_1 : 0.95, θ_2: 0.26. Additionally, the batch size is set to 4, and both pseudo and triplet losses are used with a gain of 0.1. During training for WBC detection, the triplet loss is utilized for 40 epochs, while the other loss functions are applied for the entire 50 epochs. In the case of detection with attribute prediction, the triplet loss is activated for 10 epochs, while the others for 30.

WBC Detection: We compare the proposed SLA-Det with recent state-of-the-art sparse learning methods [24], [27]. The results in Table V demonstrate that our approach significantly outperforms existing methods, achieving notably higher mAP. The superior results of our approach are due to effectively utilizing annotated information through labeled region masking, strategic use of pseudo-labels, and integrating cell similarity.

Ablation Studies: Our method contains several components.

TABLE V
COMPARATIVE RESULTS ON *Sparse-Leukemia-Attri*

Method	mAP$@_{50-95}$	mAP$@_{50}$
SparseDet [24]	6.1	10.4
Co-mining [27]	8.6	15.3
SLA-Det (Ours)	**27.2**	**45.1**

We analyze the components of the SLA-Det approach for WBC detection to verify their effectiveness. Results in Table

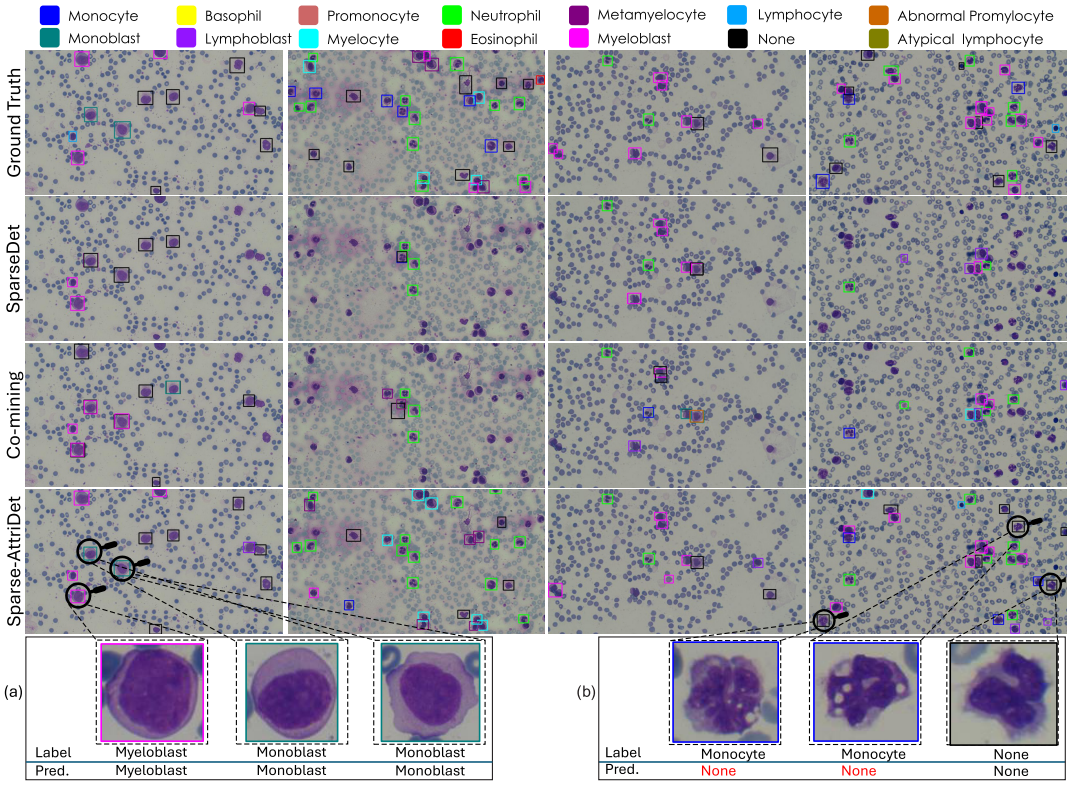


Fig. 6. Qualitative comparisons of different approaches. A zoomed-in view of cells are shown in (a) and (b)

VI show that each component contributes to the final accuracy, with LR loss improving detection mAP by 3.1%. PL loss achieves an additional 0.7% improvement and applying triplet loss, results in a significant 3.6% increase. Our results demonstrate that each component of our approach contributes effectively. We have also tried different straightforward

TABLE VI
ABLATION RESULTS OF SLA-DET ON *Sparse-Leukemia-Attri*

Labeled-Region	Pseudo-label	Triplet	mAP@$_{50-95}$	mAP@$_{50}$
✔	✘	✘	24.9	40.8
✔	✔	✘	25.1	41.5
✔	✔	✔	**27.2**	**45.1**

methods to address the sparse dataset problem. In the first experiment, the model is trained on cropped, annotated patches from *Sparse-Leukemia-Attri*, while during testing, each image is divided into overlapping patches. In the second experiment, the model is trained on the full image with unannotated regions masked out, and the entire image is used during testing. Both experiments yielded $mAP@_{50}$ scores of 37.7% and 37.6%, respectively, which are noticeably lower than the 45.1% mAP@$_{50}$ achieved by our proposed approach.

WBC detection with attribute prediction : Table. VII presents the results of our SLA-Det model for various attribute predictions. In this table, SLA-Det* denotes the AttriHead with only the LR loss. SLA-Det** includes both LR and PL losses with the AttriHead, representing the second phase of the ablation study. Finally, SLA-Det*** shows the results from the final phase of the ablation study, incorporating all losses (LR+PL+Triplet) with the AttriHead.

TABLE VII
ABLATION STUDY RESULTS: F1 SCORE OF SLA-DET ON *Sparse-Leukemia-Attri*. ATTRIBUTES ARE THE SAME AS IN TABLE III

Method	NC	NS	N	C	CB	CV	mAP@$_{50-95}$
SLA-Det*	**71.2**	**98.0**	**50.9**	**87.8**	**84.4**	**14.0**	27.6
SLA-Det**	**71.3**	**97.7**	**51.3**	**87.7**	**84.3**	**10.8**	28.1
SLA-Det***	**70.9**	**97.7**	**51.3**	**87.9**	**84.3**	**12.1**	**29.0**

Qualitative Results: The qualitative comparison of the proposed approach and existing sparse object detectors [24], [27] is shown in Figure 6. The results show that the proposed method's localization and classification capability significantly outperforms the competitive approaches. Figure 6 (a) and Figure 6 (b) show the zoomed-in version of some of the cells. In Figure 6 (a) all three cells belong to the same myeloid lineage and exhibit nearly identical morphological features. The primary difference between myeloblasts and monoblasts is the nucleus's position within the cytoplasm: monoblasts have a nucleus surrounded by a more abundant and vacuolated cytoplasm, while myeloblasts have a more distinct nuclear-cytoplasmic boundary. Despite these minor differences, the detector effectively identifies both cell types Figure 6 (b) demonstrates a failure case of our approach, where the shapes of the cell nucleus and cytoplasm are highly irregular. The first two cells are labeled as 'Monocyte' and the third cell is labeled as 'None'. We believe that the minor staining effects might be the reason to confuse the detector, potentially resulting in incorrect classifications.

VI. Conclusion

We have presented a large-scale WBC dataset named 'LeukemiaAttri+', that consists of two subsets *LeukemiaAttri* and *Sparse-Leukemia-Attri*. The *LeukemiaAttri* contains 12 subsets by using two different quality microscopes with multiple cameras at different resolutions (10x, 40x, and 100x). The collected dataset contains 14 types of WBC-level localization with their seven distinct morphological attributes. Based on morphological information, we have provided an AttriDet method to detect the WBC type with its morphological attributes. The *Sparse-Leukemia-Attri* is introduced to tackle the real-world challenges of sparse annotation in microscopy images. To establish the baseline solutions, we propose SLA-Det, which effectively utilizes labeled and unlabeled region information. We believe that the presented dataset and methods will advance research in explainable, robust, and generalized leukemia detection. Future directions may include WBCs detection alongwith with attribute detection using weakly supervised approaches.

References

[1] F. Bray *et al.*, "Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 74, no. 3, pp. 229–263, 2024.

[2] B. Rugwizangoga *et al.*, "Experience and perception of patients and healthcare professionals on acute leukemia in rwanda: A qualitative study," *Cancer Management and Research*, pp. 1923–1934, 2022.

[3] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in *CIMSA*, pp. 96–101, IEEE, 2005.

[4] R. D. Labati *et al.*, "All-idb: The acute lymphoblastic leukemia image database for image processing," in *IEEEICIP*, IEEE, 2011.

[5] S. H. Rezatofighi and H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," *Computerized Medical Imaging and Graphics*, vol. 35, no. 4, pp. 333–343, 2011.

[6] C. Matek *et al.*, "A single-cell morphological dataset of leukocytes from aml patients and non-malignant controls," *The Cancer Imaging Archive (TCIA)[Internet]*, 2019.

[7] Z. M. Kouzehkanan *et al.*, "A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm," *Scientific reports*, vol. 12, no. 1, p. 1123, 2022.

[8] A. Bodzas, P. Kodytek, and J. Zidek, "A high-resolution large-scale dataset of pathological and normal white blood cells," *Scientific Data*, vol. 10, no. 1, p. 466, 2023.

[9] Z. Gao *et al.*, "Childhood leukemia classification via information bottleneck enhanced hierarchical multi-instance learning," *IEEE TMI*, vol. 42, no. 8, 2023.

[10] A. Tomczak, S. Ilic, G. Marquardt, T. Engel, N. Navab, and S. Albarqouni, "Digital staining of white blood cells with confidence estimation," *IEEE TMI*, 2023.

[11] A. Rehman, T. Meraj, A. M. Minhas, A. Imran, M. Ali, and W. Sultani, "A large-scale multi domain leukemia dataset for the white blood cells detection with morphological attributes for explainability," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 553–563, Springer, 2024.

[12] C. Jung *et al.*, "W-net: a cnn-based architecture for white blood cells image classification," *arXiv preprint arXiv:1910.01091*, 2019.

[13] R. Singh, A. Sharma, N. Sharma, and R. Gupta, "Impact of adam, adadelta, sgd on cnn for white blood cell classification," in *2023 ICSSIT*, pp. 1702–1709, 2023.

[14] H. Wu, C. Lin, J. Liu, Y. Song, Z. Wen, and J. Qin, "Feature masking on non-overlapping regions for detecting dense cells in blood smear image," *IEEE TMI*, vol. 42, 2023.

[15] P. Huang *et al.*, "Attention-aware residual network based manifold learning for white blood cells classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 4, pp. 1206–1214, 2020.

[16] O. Saidani *et al.*, "White blood cells classification using multi-fold preprocessing and optimized cnn model," *Scientific Reports*, vol. 14, no. 1, p. 3570, 2024.

[17] T. Thanh *et al.*, "Leukemia blood cell image classification using convolutional neural network," *International journal of computer theory and engineering*, vol. 10, no. 2, 2018.

[18] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," *IEEE access*, vol. 6, pp. 36188–36197, 2018.

[19] P. Mooney, "Blood cell images." `https://www.kaggle.com/datasets/paultimothymooney/blood-cells`, April 2018. Kaggle.

[20] Z. Han *et al.*, "One-stage and lightweight cnn detection approach with attention: Application to wbc detection of microscopic images," *Computers in Biology and Medicine*, vol. 154, 2023.

[21] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.

[22] P. Manescu *et al.*, "Detection of acute promyelocytic leukemia in peripheral blood and bone marrow with annotation-free deep learning," *Scientific Reports*, vol. 13, no. 1, p. 2562, 2023.

[23] S. Tsutsui, W. Pang, and B. Wen, "Wbcatt: A white blood cell dataset annotated with detailed morphological attributes," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[24] S. Suri, S. Rambhatla, R. Chellappa, and A. Shrivastava, "Sparsedet: Improving sparsely annotated object detection with pseudo-positive mining," in *IEEE ICCV*, pp. 6770–6781, 2023.

[25] Z. Wu, N. Bodla, B. Singh, M. Najibi, R. Chellappa, and L. S. Davis, "Soft sampling for robust object detection," *arXiv preprint arXiv:1806.06986*, 2018.

[26] Y. Niitani, T. Akiba, T. Kerola, T. Ogawa, S. Sano, and S. Suzuki, "Sampling techniques for large-scale object detection from sparsely annotated objects," in *CVPR*, 2019.

[27] T. Wang, T. Yang, J. Cao, and X. Zhang, "Co-mining: Self-supervised learning for sparsely annotated object detection," in *AAAI*, vol. 35, pp. 2800–2808, 2021.

[28] A. Acevedo and A. Merino, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems," *Data in brief*, vol. 30, 2020.

[29] W. Sultani, W. Nawaz, S. Javed, M. S. Danish, A. Saadia, and M. Ali, "Towards low-cost and efficient malaria detection," in *CVPR*, pp. 20655–20664, IEEE, 2022.

[30] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[31] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, 1981.

[32] Ultralytics, "YOLOv5: A state-of-the-art real-time object detection system." `https://docs.ultralytics.com`, 2021. Accessed: 7 Jan, 2024.

[33] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *CVPR*, 2019.

[34] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, L. Li, C. Yuan, *et al.*, "Sparse r-cnn: End-to-end object detection with learnable proposals," in *CVPR*, 2021.

[35] H. Zhang, F. Li, S. Liu, L. Zhang, H. Su, J. Zhu, L. Ni, and H. Shum, "Dino: Detr with improved denoising anchor boxes for end-to-end object detection. arxiv 2022," *arXiv preprint arXiv:2203.03605*, 2022.

[36] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[37] T. Ridnik *et al.*, "Asymmetric loss for multi-label classification," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 82–91, 2021.

[38] S. Qiao, L.-C. Chen, and A. Yuille, "Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution," in *CVPR*, pp. 10213–10224, 2021.

[39] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, pp. 2117–2125, 2017.

[40] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *CVPR*, pp. 10781–10790, 2020.

[41] Y. Grandvalet and Y. Bengio, "Semi-supervised learning by entropy minimization," *Advances in neural information processing systems*, vol. 17, 2004.

[42] M. Keser, G. Schwalbe, A. Nowzad, and A. Knoll, "Interpretable model-agnostic plausibility verification for 2d object detectors using domain-invariant concept bottleneck models," in *CVPR*, 2023.

[43] M. L. Mekhalfi, D. Boscaini, and F. Poiesi, "Detect, Augment, Compose, and Adapt: Four Steps for Unsupervised Domain Adaptation in Object Detection," in *BMVC*, 2023.