Leveraging Sparse Annotations for Leukemia Diagnosis on the Large Leukemia Dataset

Abdul Rehman[a], Talha Meraj[a], Aiman Mahmood Minhas[b], Ayisha Imran[b], Mohsen Ali[a], Waqas Sultani[a], Mubarak Shah[c]

[a]*Intelligent Machine Lab, Information Technology University, Lahore, Pakistan,*
[b]*Department of Hematology, Chughtai Lab , Lahore, Pakistan,*
[c]*Center for Research in Computer Vision, University of Central Florida, United States,*

Abstract

Leukemia is 10th most frequently diagnosed cancer and one of the leading causes of cancer-related deaths worldwide. Realistic analysis of Leukemia requires White Blook Cells (WBC) localization, classification, and morphological assessment. Despite deep learning advances in medical imaging, leukemia analysis lacks a large, diverse multi-task dataset, while existing small datasets lack domain diversity, limiting real-world applicability. To overcome dataset challenges, we present a large-scale WBC dataset named 'Large Leukemia Dataset' (LLD) and novel methods for detecting WBC with their attributes. Our contribution here is threefold. First, we present a large-scale Leukemia dataset collected through Peripheral Blood Films (PBF) from several patients, through multiple microscopes, multi-cameras, and multi-magnification. To enhance diagnosis explainability and medical expert acceptance, each leukemia cell is annotated at 100x with 7 morphological attributes, ranging from Cell Size to Nuclear Shape. Secondly, we propose a multi-task model that not only detects WBCs but also predicts their attributes, providing an interpretable and clinically meaningful solution. Third, we propose a method for WBC detection with attribute analysis using sparse annotations. This approach reduces the annotation burden on hematologists, requiring them to mark only a small area within the field of view. Our method enables the model to leverage the entire field of view rather than just the annotated regions, enhancing learning efficiency and diagnostic accuracy. From diagnosis explainability to overcoming domain-shift challenges, presented datasets could be used for many challenging aspects of microscopic image analysis. The datasets, code, and demo are available at: https://im.itu.edu.pk/sparse-

leukemiaattri/.

Keywords: Explainable diagnosis, Blood cancer, leukemia, attribute, spare annotation, multi-task learning

1. Introduction

Leukemia is among the leading causes of cancer-related deaths worldwide, with an estimated 487,000 new cases and 305,000 deaths reported in 2022 [1]. This disease disproportionately affects people under the age of 40, with children being particularly impacted [2]. 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 the diagnosis and prognosis of leukemia, many remote regions in underdeveloped countries experienced higher mortality rates [3]. 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 vital to overcome the low accessibility of experts. Such a system should be able to work across multiple microscopes, and digitization scenarios, and to augment hematologists, it should provide explainable information to hematologists. Unfortunately, existing leukemia datasets [4, 5, 6, 7, 8, 9] suffer from many limitations, including the absence of localization annotations of white blood cells, the lack of associated morphological attributes, and a limited number of images [10]. Furthermore, there is no established multi-task learning (MTL) [11, 12, 13] approach that simultaneously detects white blood cells and predicts their morphological attributes.

In addition to these limitations, these supervised algorithms ([14, 15, 16, 17, 18]) require every cell in each image to be annotated, adding a significant manual burden. In the practical scenario, a hematologist uses only a small portion of the acquired image to perform the required analysis and annotation. Thus majority of the cells remain unannotated. On the other hand, requiring annotation of complete blood films not only increases time and cost

but also limits the number of blood films that could be annotated[1]. Employing existing object detection strategies over sparsely annotated datasets results in subpar performance since these methods are unable to use unlabeled parts of slides.Complete annotation of a blood film in the monolayer takes around 200 minutes—eight times longer than the 25 minutes needed for sparse annotation.

To overcome dataset challenges, we present a large-scale WBC dataset named 'Large Leukemia Dataset' (LLD), that consists of two subsets 'LeukemiaAttri'[2], capturing cell localization, classification, and morphological attribute prediction tasks; and 'Sparse-LeukemiaAttri', for the object-detection and attribute prediction based on sparsely annotated dataset task. An important highlight of the *LeukemiaAttri* is that it has been collected through two microscopes from different price ranges, one high-cost (HCM) and one low-cost (LCM), 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. In this collection, using high-resolution HCM (100x), experienced hematologists annotated 10.3k WBC of 14 types, having 55k morphological labels from 2.4k images of several PBS leukemia patients. The data set collection procedure is illustrated in Fig.2 and further details are in section 3. In conjunction with these datasets, we propose two models named Leukemia Attribute Detector (AttriDet) and Sparse Leukemia Attribute Detector (SLA-Det). For the task of detection of white blood cells (WBCs) along with the prediction of morphological attributes, we integrated a lightweight head, named AttriHead, into the existing object detector. AttriHead employs multiscale features from the feature backbone used in object detection. This not only enhances computational efficiency but also enables the backbone to become more informative through MTL. For the sparsely-supervised object detection, we propose a novel algorithm called SLA-Det. Note that in our case on average only 20% of the image is annotated shown in Figure 1. To effectively use the unannotated region information, SLA-Det incorporates region-guided masking and pseudo-label selection based on multiple criteria, ensuring robust model performance. We believe that these extensions align our dataset and framework more with clinical practices and

[1]Note that multiple image patches are extracted at the specified resolution

[2]This subset was published in *'A Large-scale Multi Domain Leukemia Dataset for the WBC Detection with Morphological Attributes for Explainability'* at MICCAI 2024 [19], and this journal paper serves as its extension.

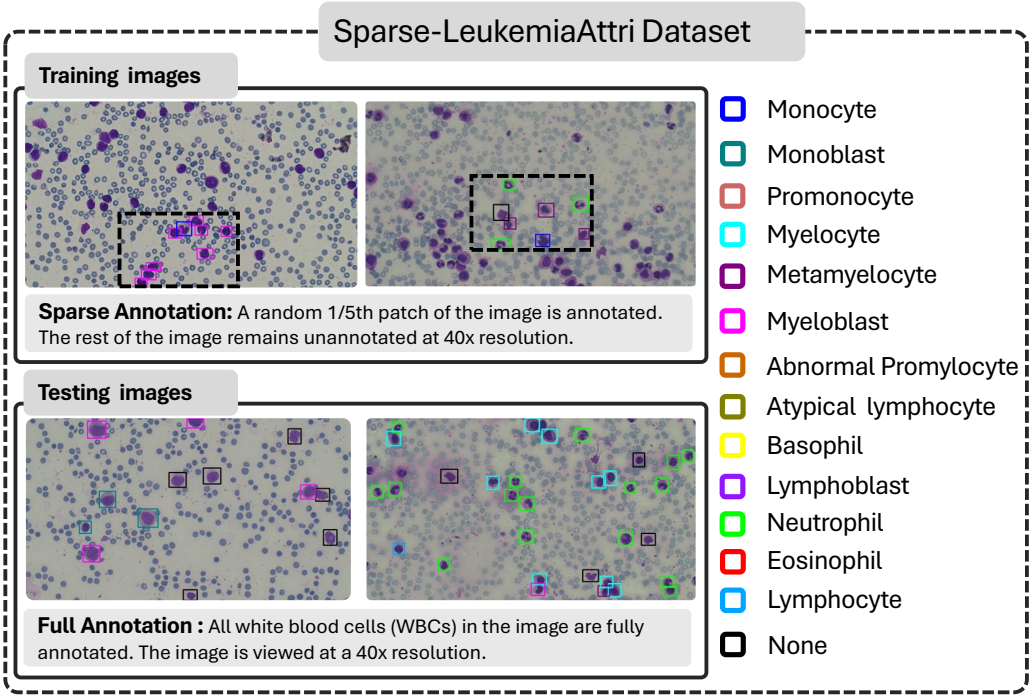


Figure 1: Sparse-LeukemiaAttri Dataset: In the training set, images are sparsely annotated, with only a random $(1/5)^{th}$ of each image labeled. For the testing set, the entire image is fully annotated to allow for comprehensive evaluation.

enhance the reliability of automated leukemia detection. Our extensive experimental results and discussion validate the proposed ideas and framework. Our contributions are given below.

- **Large Leukemia Dataset (LLD)** consists of LeukemiaAttri and Sparse-LeukemiaAttri. **LeukemiaAttri:** A large-scale, multi-domain dataset of 2.4K microscopic images captured using microscopes of varying costs, two cameras, and three resolutions (10x, 40x, 100x). Each image includes WBC localization and morphological attribute annotations for an explainable diagnosis. **Sparse-LeukemiaAttri:** A 40x FoV dataset with sparsely annotated training data (1,546 images) and fully annotated test data (185 images).
- Two detection models are proposed:
  - **AttriDet** is a fully supervised detector designed to identify WBCs

and their morphological attributes.

  - **SLA-Det** is a sparse WBC detector that enhances pseudo-label selection by incorporating problem-specific constraints, including size, objectness score, and class label entropy.

- Key results: SLA-Det processes slides 4x faster than AttriDet at 40x resolution (vs. 100x) while maintaining accuracy. SLA-Det outperforms state-of-the-art sparse detection methods by 34.7 $mAP@_{50}$, demonstrating superior efficiency and performance.

The work emphasizes domain diversity, diagnostic interpretability, and efficient sparse annotation strategies for real-world leukemia analysis.

2. 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 that highlights their relevance to WBC detection.

Supervised WBC detection: Deep learning (DL) techniques have been extensively applied to WBC classification [20, 21, 22, 23]. Huang *et al.* [24] 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.* [25] optimized a CNN model and employed multiple pre-processing techniques, such as color conversion and image histogram equalization. Similarly, the authors in [26] classified two types of WBCs from the ALL-IDB data set [26]. G Liang et al. [27] developed a combined model of convolutional neural networks (CNN) and recurrent neural networks (RNN) to classify four types of WBCs using the BCCD dataset [28].

To detect WBCs in whole microscopy images, the authors in [29] gave a lighter deep-learning CNN model and adapted the YOLOX-nano [30] baseline to identify the types of WBC. 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 [31]. For image preprocessing, they mainly applied threshold methods.

Sparsely annotated object detection: To tackle the challenge of object detection in sparsely annotated datasets, authors in [32] 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.* [33] 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. Niitani et al. [34] propose a part-aware learning approach that trains a model twice, using pseudo labels from the first model to guide the second. By leveraging spatial relationships, it selectively ignores classification loss for part categories within subject categories, improving training and performance with sparse annotations. To generate labels for sparsely annotated datasets, the co-mining approach is applied by authors of [35]. It involves constructing a Siamese detection network with two branches, each generating pseudo-label sets using a co-generation module. 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 [32, 35], 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.

3. Dataset

Leukemia comprises four main types: Acute lymphocytic leukemia (ALL), Chronic lymphocytic leukemia (CLL), Acute myeloid meukemia (AML), and Chronic myeloid meukemia (CML) [1]. These types are categorized into two major lineages: myeloid (AML and CML) and lymphoid (ALL and CLL). Each leukemia type exhibits distinct abnormalities, including increased myelo-

blasts in CML, atypical lymphocytes in CLL, elevated myeloblasts in AML, and excessive lymphoblasts in ALL. However, as shown in Table 1, most existing white blood cell (WBC) datasets were collected using a single microscope, lacked bounding boxes and morphological details, and often included either only healthy individuals or patients with a single leukemia type.

Table 1: Comparison of the proposed dataset with existing leukemia datasets.

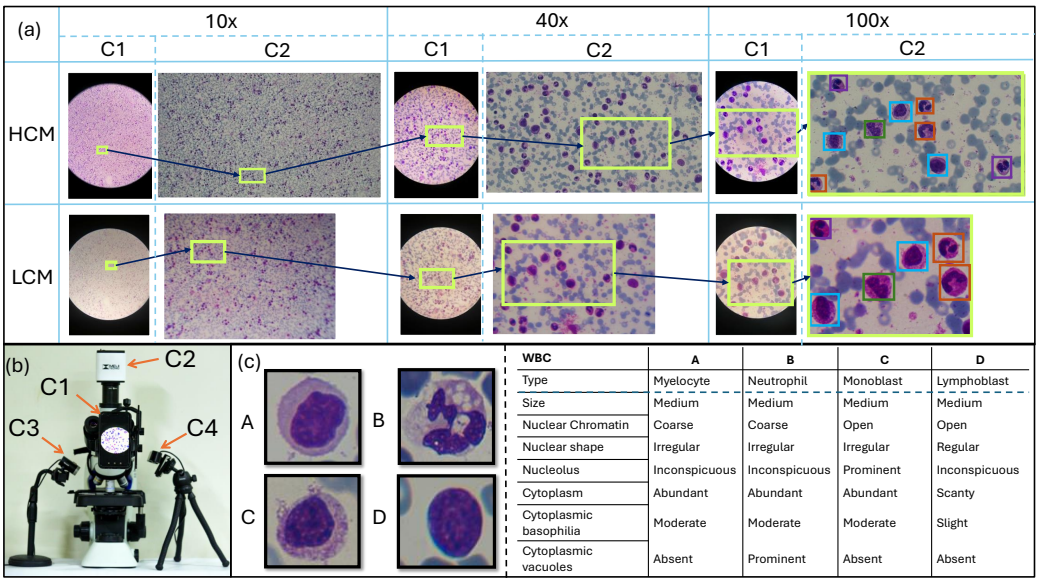
Dataset	Type	Multi. Micro.	Multi. Cells in image	BBX	Multi. Res.	No. of WBC's	WBC Classes	Morphology
IDB [4]	ALL	✘	✔	✔	✘	510 (LB)	2	✘
IDB2 [5]	ALL	✘	✘	✘	✘	260	2	✘
LISC [6]	Normal	✘	✘	✘	✘	250	5	✘
Munich[7]	AML	✘	✘	✘	✘	18,365	15	✘
Raabin [8]	Normal	✘	✔	✔	✘	17,965	5	✘
HRLS [9]	Multi.	✘	✔	✘	✘	16,027	9	✘
WBCAtt [36]	Normal	✘	✘	✘	✘	10,298	5	✔
LeukemiaAttri [3]	Multi.	✔	✔	✔	✔	**88,294**	14	✔

3.1. Limitations of existing WBC datasets

The details of publicly available datasets on WBC and leukemia are shown in Table 1, which highlights various limitations such as: 1) The ALL-IDB dataset [4] contains only 510 WBCs from patients with Acute Lymphocytic Leukemia (ALL) and includes only cell centroid information. 2) In version 2 of the ALL IDB [5], each image contains only a single WBC, with the cell's status labeled accordingly. 3) The LISC [6] 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 [7] contains single-cell images that come under the umbrella of AML. 5) Only Raabin [8], out of all the aforementioned existing datasets, has complete bounding box-level annotation with single-cell images, but the dataset consists entirely of normal individuals, with over 60% of the images containing only a single cell in the FoV. Furthermore, the Raabin-WBC collection contains images from two separate cameras and two different microscopes, without maintaining any correspondence across microscopes. 6) The HRLS [9] dataset is captured using a single microscope and covers only two types of Leukemia and nine types of WBCs. 7) WBCAtt dataset [36] is the improved version of [37],

[3]The details LeukemiaAttri is explained in Table 2

where annotations of attributes are added for each WBC. The dataset [37] lacks the abnormal WBCs and is captured using a single microscope only.



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

Figure 2: a) Illustrates the image acquisition procedure utilizing standard mobile camera (C1), premium camera for high-cost microscopy (HCM), and mid-range camera (C2) for low-cost microscopy (LCM) at various resolutions (100x, 40x, and 10x). The images are obtained with both high and low-cost microscopes. b) Displays our microscope set up with four cameras C1 to C4. C1 and C2 capture the field of view of the microscope lens, and C3 and C4 capture stage scale. c) This table presents various types of white blood cells (WBC) and their distinctive morphological features.

3.2. Our Dataset Collection Procedure

To understand leukemia prognosis, we consulted healthcare professionals to finalize key WBC types and their morphology. To improve deep learning across modalities, we propose capturing images of four leukemia PBF types from multiple patients at various magnifications using available resources. Capturing images at varying resolutions enhances deep learning generalization, enabling low-cost, automated leukemia diagnostics.

In this dataset collection, PBFs are sourced from a diagnostic lab with patient consent and include annotations from a hematologist in the monolayer area. Images are captured using two microscopes: the high-cost Olympus CX23 and the low-cost XSZ-107BN, along with three cameras: the HD1500T

(HCM), ZZCAT 5MP (LCM), and the Honor 9x Lite mobile camera. The high-end camera (ZZCAT 5MP) is 47 times more expensive than the low-end camera (XSZ-107BN), while the HCM is 17 times more costly than the LCM. Although the LCM is more affordable, it sacrifices visual detail and has a smaller field of view. Note that capturing the same WBC across multiple microscopes and resolutions is highly challenging due to several factors: (a) Locating the same WBC patch on the peripheral blood film (PBF) across different microscopes and resolutions is inherently difficult [38], (b) Microscope stage scales are not calibrated uniformly across different microscopes, (c) Even within the same microscope, poor alignment leads to resolution inconsistencies, and (d) Viewing a patch at 100x requires immersion oil, whereas 10x and 40x provide better visibility without it. This necessitates repeatedly adding and removing oil when capturing the same patch at multiple resolutions.

We began the capturing process by setting the field of view (FoV) at 10x with approximately 20% overlap while keeping a fixed x-axis microscope stage scale. We then replicated the same procedure for the 40x magnification at an identical scale. At 100x magnification, we captured the corresponding FoV without overlap, ensuring each WBC was distinctly represented. To achieve this, we used an in-house developed software solution that enables simultaneous FoV image capture from mobile (C1), microscope-specific cameras (C2), and stage-capturing cameras (C3 and C4). For cross-microscope comparison, we first calibrated the scale stages of both microscopes and then repeated the process for the LCM.

3.3. LeukemiaAttri dataset

To overcome the above-mentioned dataset limitations and provide a benchmark for training deep learning models, we present a large-scale dataset called LeukemiaAttri dataset. The LeukemiaAttri dataset comprises 14 types of WBCs. The details of each type in a subset are presented in Table 3, where the "None" category includes artifacts, bare cells, and cells that were challenging for hematologists to identify. In total, the data set contains 28.9K images (2.4K images $\times$ 2 microscopes $\times$ 3 resolutions $\times$ 2 camera types), representing 12 subsets of images captured using low- and high-cost microscopes at three magnification levels. The detail of the LeukemiaAttri dataset is explained in Table 2.

Table 2: LeukemiaAttri dataset details with its multi-domain subsets

Sr. #	Dataset Name	Microscope	Resolution	Camera	Image Count	WBC Count	Artifacts
1	HCM_100x_C1	OlympusCX23	100x	Honor 9X lite	2.4k	7.8k	2.5k
2	HCM_100x_C2	OlympusCX23	100x	HD1500T	2.4k	7.8k	2.5k
3	HCM_40x_C1	OlympusCX23	40x	Honor 9X lite	2.4k	7.8k	2.5k
4	HCM_40x_C2	OlympusCX23	40x	HD1500T	2.4k	7.8k	2.5k
5	HCM_10x_C1	OlympusCX23	10x	Honor 9X lite	2.4k	7.8k	2.5k
6	HCM_10x_C2	OlympusCX23	10x	HD1500T	2.4k	7.8k	2.5k
7	LCM_100x_C1	XSZ_107BN	100x	Honor 9X lite	2.4k	4.8k	1.1k
8	LCM_100x_C2	XSZ_107BN	100x	ZZCAT-5MP	2.4k	4.8k	1.1k
9	LCM_40x_C1	XSZ_107BN	40x	Honor 9X lite	2.4k	7.8k	2.5k
10	LCM_40x_C2	XSZ_107BN	40x	ZZCAT-5MP	2.4k	7.8k	2.5k
11	LCM_10x_C1	XSZ_107BN	10x	Honor 9X lite	2.4k	7.8k	2.5k
12	LCM_10x_C2	XSZ_107BN	10x	ZZCAT-5MP	2.4k	7.8k	2.5k
Sparse-LeukemiaAttri							
13	HCM_40x_C2_Sparse	OlympusCX23	40x	HD1500T	1.7k	7.6k	2.0k

Table 3: WBC types and their distribution in a single subset of the LeukemiaAttri (**A**) and Sparse-LeukemiaAttri (**B**) datasets.

Sr. #	Type of cell	Count in **A**	Count in **B**	Sr. #	Type of cell	Count in **A**	Count in **B**
1	Monocyte	205	189	8	Monoblast	484	332
2	Basophil	19	18	9	Lymphoblast	1722	1588
3	Promonocyte	346	335	10	Myelocyte	304	329
4	Neutrophil	1293	1153	11	Eosinophil	85	95
5	Metamyelocyte	101	133	12	Myeloblast	2157	2029
6	Lymphocyte	322	282	13	Atypical lymphocyte	419	420
7	Abnormal promyelocyte	276	298	14	None	2560	2375

3.3.1. Morphological attributes:

To improve prognostic analysis, hematologists identified 14 distinct types of WBCs. The morphological characteristics of several types of WBC are illustrated in Figure 2 (c). For annotating the WBC types and their morphological features, hematologists meticulously reviewed all subsets of the captured images and selected the subset with the best-quality images that offered clear structural details. This selected subset (HCM_100x_C2) , which was captured using the High-Cost Microscope (**HCM**) at **100x** magnification with the HD1500T camera (**C2**) was used for the annotation process.
To ensure accuracy, two hematologists collaborated in annotating the cells. Figure 2 (c) shows examples of certain WBC types and their morphological attributes, such as monoblasts, monocytes, and myelocytes, (A, B, C in the

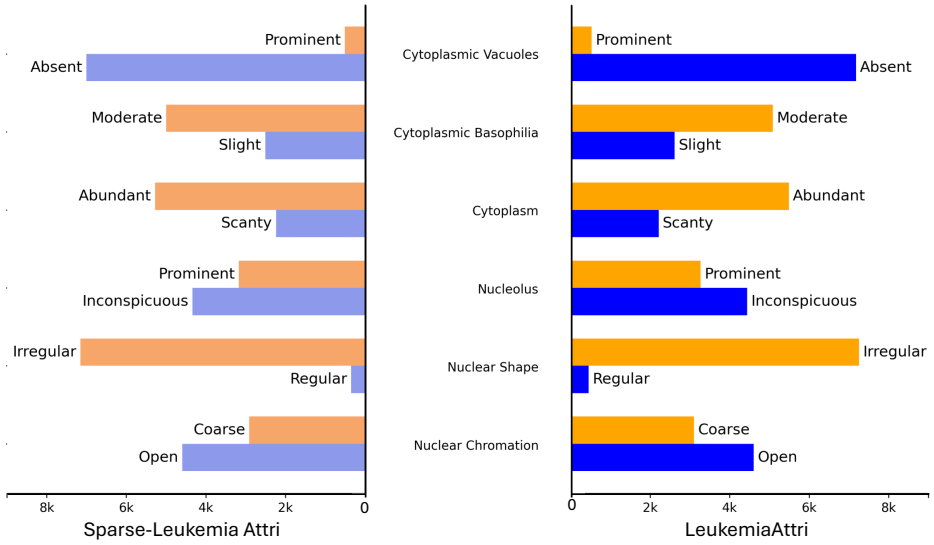


Figure 3: Attribute classes distribution of a subset from LeukemiaAttri and Sparse leukemiaAttri datasets. The distribution shows the class imbalance in the "Nuclear shape and Cytoplasmic vacuoles is much higher than the others.

Figure 2(c)) which exhibit similar morphological features due to their shared myeloid lineage. However, variations are observed, especially in the presence of cytoplasmic vacuoles. On the other hand, lymphoblasts (Figure 2(c): D), which belong to the lymphoid lineage, exhibit distinct morphological characteristics compared to the myeloid lineage. Once detailed annotations of WBC types and their attributes were obtained at 100x magnification using the HCM, we transferred these annotations to images captured at different resolutions and with the LCM using homography techniques [39, 40]. The transferred annotations were manually reviewed, and any missing or incorrectly localized annotations were re-annotated. The specific counts of WBC types and their corresponding attributes for the source subset are provided in Table 3

3.4. Sparse LeukemiaAttri dataset

To ensure the accuracy of morphological attributes, the LeukemiaAttri dataset was collected at 100x magnification and then the corresponding patches from 40x and 10x were extracted while ignoring the complete FOV of 40x and 10x. To make efficient use of available data and to ensure practical efficacy (since 40x is the gold standard for Leukemia), in the Sparse

LeukemiaAttri dataset, the complete FOV of 40x should be considered. To address the above limitation of the LeukemiaAttri dataset, we extend the HCM_40X_C2 (see Table 2) subset to include a complete FoV. This new dataset is split up into a train set (1546 images) and a test-set (185 images). For the training set, the annotations are borrowed from the LeukemiaAttri dataset and mapped to 40x images, thus resulting in the sparse annotations and named as *Sparse-LeukemiaAttri*. The ratio of the labeled patch to the entire field of view (FoV) at 40x, in terms of size, is 1:5. However, for the test set (curated at 40x resolution), each patch of the image is annotated at 100x (for better annotation quality) resolution and these annotations are transferred to a complete FOV of 40x test set. Examples of train and test subsets are shown in Figure 1. Collectively, with a new extension, we have 13 subsets now as shown in Table 2.
Our *Sparse-LeukemiaAttri* 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.

4. Proposed Multi Task Detectors

Our goal is to detect WBC, along with their morphological attributes, from PBF images to enable a real-time, explainable diagnosis of leukemia. To avoid large computational cost, instead of separate network we extract features from the backbone of the predicted WBC and estimate the attributes of the WBC through multi-task learning. Furthermore, to adhere with the real-word scenario, where hematologists analyze the PBF on 40x for efficiency, proposed WBC detector and attribute predictor is trained on complete FOV of 40x resolution. Since data was annotated on 100x, it maps to only small portion in 40x, resulting in sparse annotation. Therefore, we design self-learning mechanism train Multi Task object detector in a sparse setting.

4.1. 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 $I_i \in \mathbb{R}^{H \times W}$ represents 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 [17] for our approach, since our experiments indicated (Table 4) it to have higher mAP in comparison with other recent SOTA object detectors including FCOS [15], Spare-Faster-RCNN [14] and transformer-based DINO[16].

4.2. AttriDet: WBC detection with attributes

Our objective is to augment existing object detectors with morphological attribute prediction modules. Let $\mathcal{B}$ be the feature extractor part of the object detector $\mathcal{F}$ shown in Figure 4. 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 $S_j \in \mathbb{R}^{H_j \times W_j \times C_j}$. 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 is 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 [41] to equalize the dimensions and properly align the extracted feature. Let $\mathcal{ROI}$ denote the ROIAlign operation.

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

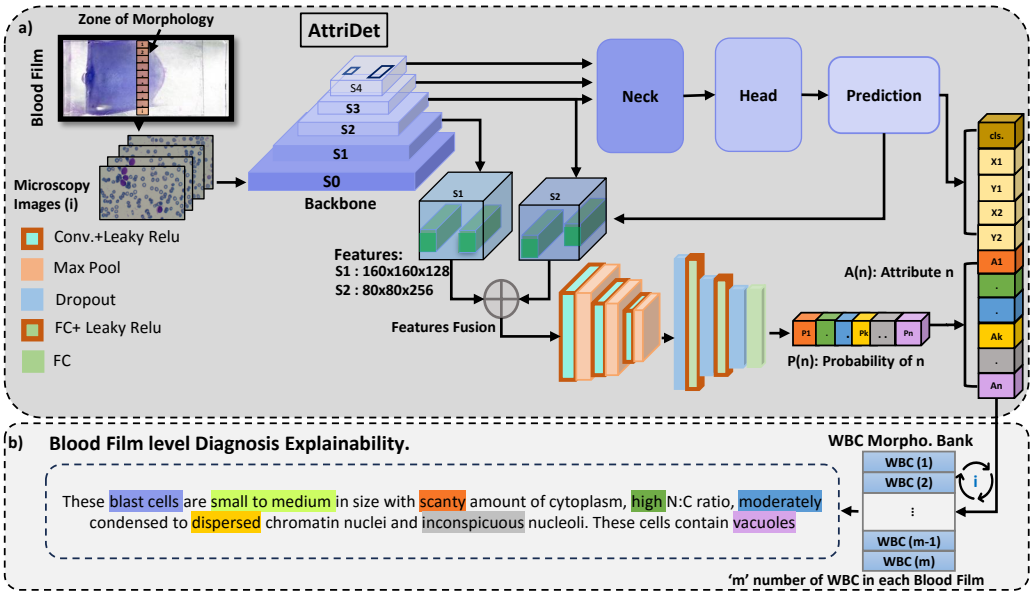


Figure 4: Architecture of Attri-Det: The proposed framework for white blood cell (WBC) and morphological attribute detection processes input images using a backbone network to extract features. The WBC detector is trained with standard detection losses, while the attribute head uses asymmetric multilabel classification loss. b) Predictions from each image are stored in a morphology bank, updating the 'Masked Text' (shown in the different colors in the figure) with the predominant WBC type and morphology for a comprehensive blood film-level description.

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 shown in Figure 4 (a). 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 [42]. Once trained, the AttriHead predicts the six-dimensional vectors $\hat{\mathcal{M}}^{1\times 6}$, where each dimension represents the state of the attribute. The details of the morphological attributes are provided in Section 3.3.1. For the MTL, in addition to the AttriHead, we also train the YOLOv5 for WBC detection, utilizing the following loss

functions with the standard settings of the detector:

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

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.

Attridet collectively detects cells and predicts their morphology, providing diagnostic reasoning. For each PBF image, these predictions are recorded in a morphology bank, and a blood film-level description is updated in the 'Masked Text' based on the most frequently occurring WBC type and morphology, as recommended by hematologists. An example of an updated text-based description is shown in Fig. 4 (b). We presented the text generated from AttriDet's predictions to hematologists, who recognized its potential as a valuable second-opinion tool for leukemia prognosis. Furthermore, AttriDet not only predicts cell-associated attributes but also improves the detector's performance, raising the mAP from 26.3 to 28.2, as shown in Table 5.

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

Due to annotation cost, images in the training set of *Sparse-LeukemiaAttri* are sparsely annotated. 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 [32] and Co-mining [35], 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

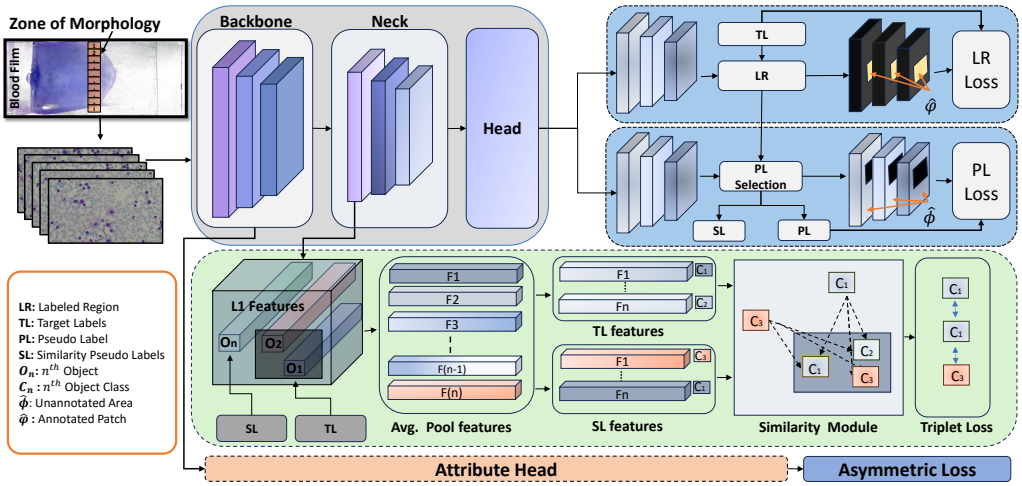


Figure 5: Architecture of SLA-Det: The proposed framework detects sparsely annotated white blood cells (WBCs) and morphological attributes in 40× images, using a backbone for feature extraction and an AttriHead module. Labeled regions (LR) optimize predictions via labeled region loss against ground-truth labels (TL), while unlabeled regions employ pseudo-labels (PL, high confidence) and similarity pseudo labels (SL) with triplet loss to align class-based features. Training combines labeled regions, pseudo-labels, morphology-attributes, and triplet losses, enabling robust robust learning for accurate detection in sparse annotation scenarios.

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) is proposed.

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 a fully annotated region. Note that this region is about 20% of the image (as shown in Figure 1). 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 $\hat{\phi}_i = \{\text{set of all } \hat{\mathbf{z}}_{i,k} \text{ that do not lie inside } P_i\}$ and $\hat{\varphi}_i = \{\text{set of all } \hat{\mathbf{z}}_{i,k} \text{ that lie inside } P_i\}$.

4.3.1. Training on Annotated Regions

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 ([43, 44, 45]) 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 $P^v_{i,\text{norm}}$. For each level, we separate the predictions, such that $\forall_v \hat{\varphi}^v_i = \{\text{set of all } \hat{\mathbf{z}}_{i,k} \text{that lie inside } P^v_i\}$.

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 $\hat{\varphi}^v_i$ 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 annotated region is Eq. (4):

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

4.3.2. Training on non-Annotated Regions

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 from the not-properly trained model are noisy and might be incorrect, we design filtering criteria to select the pseudo-labels from the predictions. First, we filter out predictions, $\mathbf{z}^*_{i,k}$, that have predicted confidence score, $p^*_{i,k}$ below $\mathbf{t}_0$. 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 ($\text{Area}(\hat{b}_{i,k}) > \text{Area}_{\text{min}}$) in the ground-truth in the training set. Thirdly, in the remaining set, for any prediction with

confidence score $\hat{p}^*_{i,k}$ above $\mathbf{t}_1$, we compute entropy, allowing us to focus on the most distinguished class probabilities [46].

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

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,

$$\begin{aligned} D_i^{pseudo} = \{\hat{\mathbf{z}}_{i,k} \in \hat{\phi}_i \mid \text{Area}(\hat{b}_{i,k}) > \text{Area}_{\min}) \ \& \\ \hat{c}_{i,k} > \mathbf{t}_1 \ \& \ \mathcal{E}_{i,k} < \mathbf{t}_2\}, \end{aligned} \tag{6}$$

where $\mathbf{t}_2$ 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 the 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 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 $\mathbf{t}_0$ indicating they might not be background but less than $\mathbf{t}_1$ (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, to learn to predict attributes (using AttriHead) along with WBC detections, we employ the following loss.

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

5. Experiments and Results

5.1. 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 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: of 0.01 to 0.001) and a 3-epoch warmup.

WBC Detection: To evaluate WBC detection performance on *LeukemiaAttri*, we perform experiments with several well-known object detectors including Sparse R-CNN [14], FCOS [15], DINO[16], and YOLOv5x [17]. 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 cameras (HD1500T for HCM and MZZCAT 5MP for LCM - C2) on 40x and 100x magnifications. The experimental results are shown in Table 4. 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.

Table 4: Object Detection baselines results ($mAP_{@50}$) on multipal subset of *LeukemiaAttri* dataset. The C2 camera represents the HD1500T camera for HCM, while the C2* camera corresponds to the ZZCAT 5MP camera for LCM. The C1 mobile camera, used for both, is Honor 9x Lite.

Methods	High cost Microscope				Low cost Microscope			
	100x		40x		100x		40x	
	C1	C2	C1	C2	C1	C2*	C1	C2*
Sparse R-CNN [14]	29.6	36.7	27.0	32.7	32.6	25.9	26.4	33.9
FCOS [15]	31.8	40.6	24.8	32.7	33.9	34.3	31.2	28.5
DINO[16]	33.8	43.7	**36.1**	36.9	34.3	**38.2**	31.4	**36.6**
YOLOv5x[17]	**38.8**	**44.2**	**36.1**	**37.3**	**39.5**	29.7	**38.1**	34.9

WBC detection with attribute prediction: Table. 5 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 [47] and AttriDet demonstrate superior performance in predicting WBC types and their

attributes. However, detecting nucleus and cytoplasmic vacuoles remains a significant challenge.

Table 5: 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 [47]	21.9	**96.2**	41.8	77.2	70.2	3.33	27.6
AttriDet [19]	**73.9**	95.9	**54.3**	**89.7**	**83.6**	**29.1**	**28.2**

Unsupervised Domain Adaptation based WBC detection: The LeukemiaAttri dataset includes 12 subsets originating from distinct domains, captured via HCM and LCM. These subsets exhibit significant domain shifts, as illustrated in Fig. 2 (a), positioning our dataset as a valuable benchmark for domain adaptation studies. It is worth noting that LCM images often suffer from poor quality, discouraging hematologists from providing annotations due to the labor-intensive and error-prone nature of the process. As a solution, UDA approaches enable the training of object detectors on high-quality, well-annotated HCM images and facilitate their application to LCM images. To establish the UDA baselines, we evaluated UDA baselines using ConfMix

Table 6: Domain adaptation results for object detection on the *LeukemiaAttri* dataset

Method	$mAP@_{50-95}$	$mAP@_{50}$
YOLOv5[17]	11.0	25.5
DACA [48]	12.6	30.2
ConfMix [14]	12.6	33.5

[14] and DACA [48] on high-resolution subsets (HCM_100x_C2 as source and LCM_100x_C2 as target). Table 6 demonstrates the YOLOv5x (source only) achieved 25.5 $mAP@_{50}$, while UDA methods improved performance to 33.5 (ConfMix) and 30.2 (DACA), highlighting the significant domain shift and complexity of our dataset.

5.2. Experiments on Sparse-LeukemiaAttri

Implementation details: For the SLA-Det implementation, we set the values as follows: $\mathbf{t}_0 : 0.70$, $\mathbf{t}_1 : 0.95$, $\mathbf{t}_2$: 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. not clearIn 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 [32, 35]. The results in Table 7 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.

Table 7: Comparative results on *Sparse-LeukemiaAttri*

Method	mAP@$_{50-95}$	mAP@$_{50}$
SparseDet[32]	6.1	10.4
Co-mining[35]	8.6	15.3
SLA-Det (Ours)	**27.2**	**45.1**

Ablation Studies: Our method contains several components. We analyze the components of the SLA-Det approach for WBC detection to verify their effectiveness. Results in Table 8 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

Table 8: Ablation results of SLA-Det on *Sparse-LeukemiaAttri*

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

straightforward methods to address the sparse dataset problem. In the first experiment, the model is trained on cropped, annotated patches from *Sparse-LeukemiaAttri*, 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.

Table 9: Ablation study results: F1 Score of SLA-Det on *Sparse-LeukemiaAttri*. Attributes are the same as in Table 5

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**

WBC detection with attribute prediction : Table. 9 presents the results of our SLA-Det model for various prediction of attributes. 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.

Qualitative Results: The qualitative comparison of the proposed approach and existing sparse object detectors [35, 32] 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.

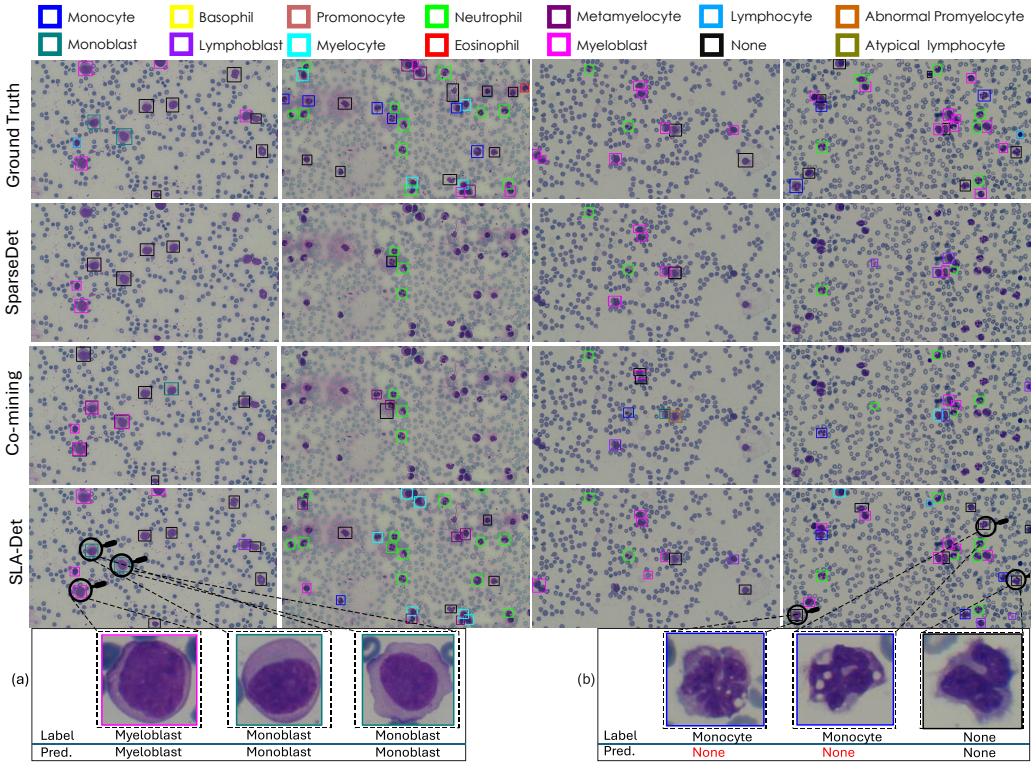


Figure 6: Qualitative comparison of different sparsely annotated approaches, including SparseDet [32] and Co-mining [35], and SLA-Det. Previous methods struggle to accurately localize all WBCs (columns 2 and 4) whereas the proposed approach effectively detects WBCs with improved classification. A zoomed-in view highlights (a) successful detection cases and (b) misclassified cases labeled as None (details are given in text).

6. Conclusion

We present a Large Leukemia Dataset (LLD) consisting of datasets named *LeukemiaAttri* and *Sparse-LeukemiaAttri*. The *LeukemiaAttri* contains 12 subsets collected by using two different quality microscopes with multiple cameras at different resolutions (10x, 40x, and 100x). There are 2.4k annotated images per resolution (in both HCM and LCM), with 10.3K cells annotated for 14 WBC types and 7 morphological properties. Based on morphological information, we have provided an AttriDet method to detect the WBC type with its morphological attributes. The *Sparse-LeukemiaAttri* 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.

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 74 (2024) 229–263.

[2] U. Creutzig, M. A. Kutny, R. Barr, R. F. Schlenk, R. C. Ribeiro, Acute myelogenous leukemia in adolescents and young adults, Pediatric blood & cancer 65 (2018) e27089.

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

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

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

[6] S. H. Rezatofighi, H. Soltanian-Zadeh, Automatic recognition of five types of white blood cells in peripheral blood, Computerized Medical Imaging and Graphics 35 (2011) 333–343.

[7] 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).

[8] 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 12 (2022) 1123.

[9] A. Bodzas, P. Kodytek, J. Zidek, A high-resolution large-scale dataset of pathological and normal white blood cells, Scientific Data 10 (2023) 466.

[10] Z. Gao, et al., Childhood leukemia classification via information bottleneck enhanced hierarchical multi-instance learning, IEEE TMI 42 (2023).

[11] Q. He, Q. Yang, H. Su, Y. Wang, Multi-task learning for segmentation and classification of breast tumors from ultrasound images, Computers in Biology and Medicine 173 (2024) 108319.

[12] M. Xu, K. Huang, X. Qi, A regional-attentive multi-task learning framework for breast ultrasound image segmentation and classification, IEEE Access 11 (2023) 5377–5392. doi:10.1109/ACCESS.2023.3236693.

[13] C. F. Koyuncu, G. N. Gunesli, R. Cetin-Atalay, C. Gunduz-Demir, Deepdistance: a multi-task deep regression model for cell detection in inverted microscopy images, Medical Image Analysis 63 (2020) 101720.

[14] 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.

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

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

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

[18] T. Wollmann, K. Rohr, Deep consensus network: Aggregating predictions to improve object detection in microscopy images, Medical Image Analysis 70 (2021) 102019.

[19] A. Rehman, T. Meraj, A. M. Minhas, A. Imran, M. Ali, 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, Springer, 2024, pp. 553–563.

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

[21] R. Singh, A. Sharma, N. Sharma, R. Gupta, Impact of adam, adadelta, sgd on cnn for white blood cell classification, in: 2023 ICSSIT, 2023, pp. 1702–1709. doi:10.1109/ICSSIT55814.2023.10061068.

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

[23] C.-W. Wang, S.-C. Huang, Y.-C. Lee, Y.-J. Shen, S.-I. Meng, J. L. Gaol, Deep learning for bone marrow cell detection and classification on whole-slide images, Medical Image Analysis 75 (2022) 102270.

[24] P. Huang, et al., Attention-aware residual network based manifold learning for white blood cells classification, IEEE Journal of Biomedical and Health Informatics 25 (2020) 1206–1214.

[25] O. Saidani, et al., White blood cells classification using multi-fold preprocessing and optimized cnn model, Scientific Reports 14 (2024) 3570.

[26] T. Thanh, et al., Leukemia blood cell image classification using convolutional neural network, International journal of computer theory and engineering 10 (2018).

[27] G. Liang, H. Hong, W. Xie, L. Zheng, Combining convolutional neural network with recursive neural network for blood cell image classification, IEEE access 6 (2018) 36188–36197.

[28] Shenggan, N. Chen, cosmicad, akshaylamba, Bccd: Blood cell count and detection, 2018. URL: `https://github.com/Shenggan/BCCD_Dataset`.

[29] 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 154 (2023).

[30] Z. Ge, S. Liu, F. Wang, Z. Li, J. Sun, Yolox: Exceeding yolo series in 2021, arXiv preprint arXiv:2107.08430 (2021).

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

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

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

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

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

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

[37] A. Acevedo, A. Merino, A dataset of microscopic peripheral blood cell images for development of automatic recognition systems, Data in brief 30 (2020).

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

[39] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2004) 91–110.

[40] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981).

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

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

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

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

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

[46] Y. Grandvalet, Y. Bengio, Semi-supervised learning by entropy minimization, Advances in neural information processing systems 17 (2004).

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

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