

A Large-scale Multi Domain Leukemia Dataset for the White Blood Cells Detection with Morphological Attributes for Explainability

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

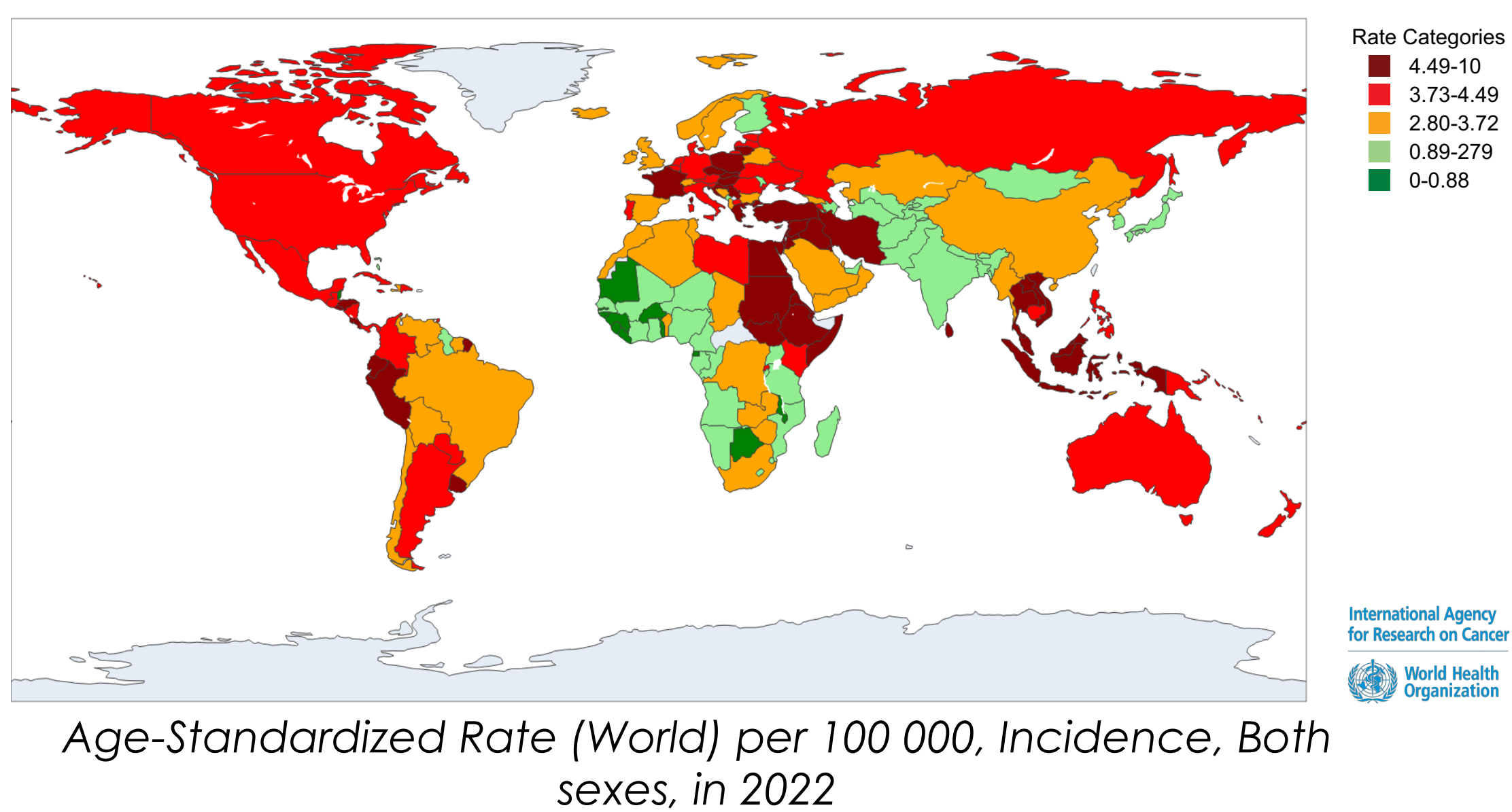
* Intelligent Machines Lab, Information Technology University, Lahore, Pakistan ** Department of Hematology, Chughtai Labs, Lahore, Pakistan

Problem Statement

Leading cause of cancer-related deaths under 39 years

10th most frequently diagnosed cancer

Frightening situation due to its quick divergence

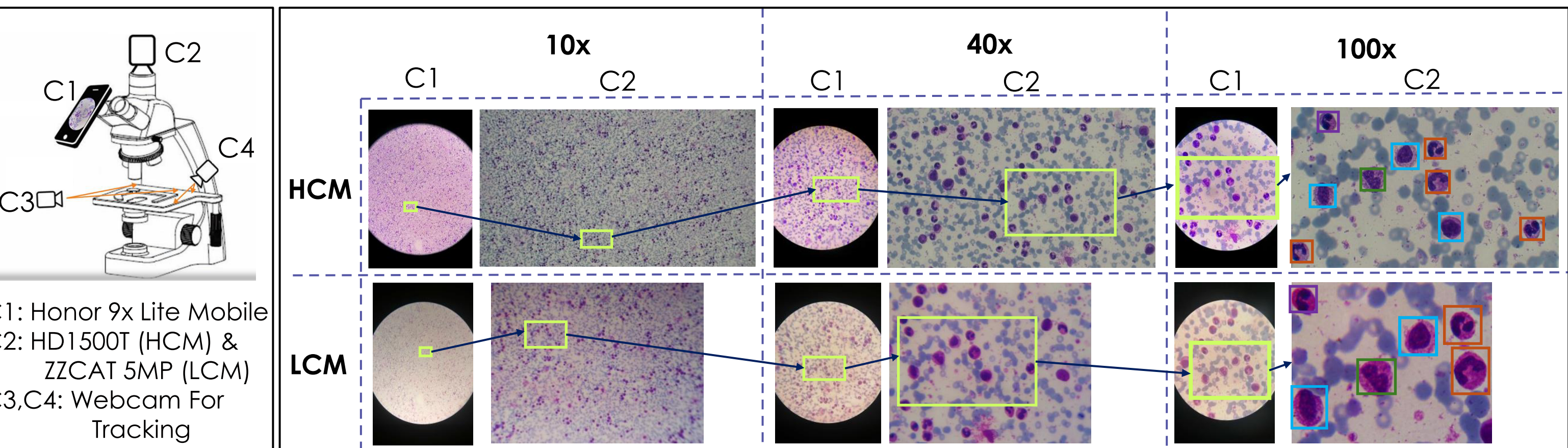


- For low cost, microscopy analysis is knowledge intensive and prone to errors.
- Automatic explainable diagnosis is hampered by lack of sufficient dataset.

Contributions

- Introducing a benchmark for WBC detection and morphological attributes to aid leukemia prognosis.
- Established WBC detection and UDA baselines to facilitate the future research
- Introducing AttriDet: a multi-headed WBC detection and attribute prediction architecture

Dataset Collection



Significance of WBC's Morphology

Importance:

Disease Diagnosis

Disease Classification

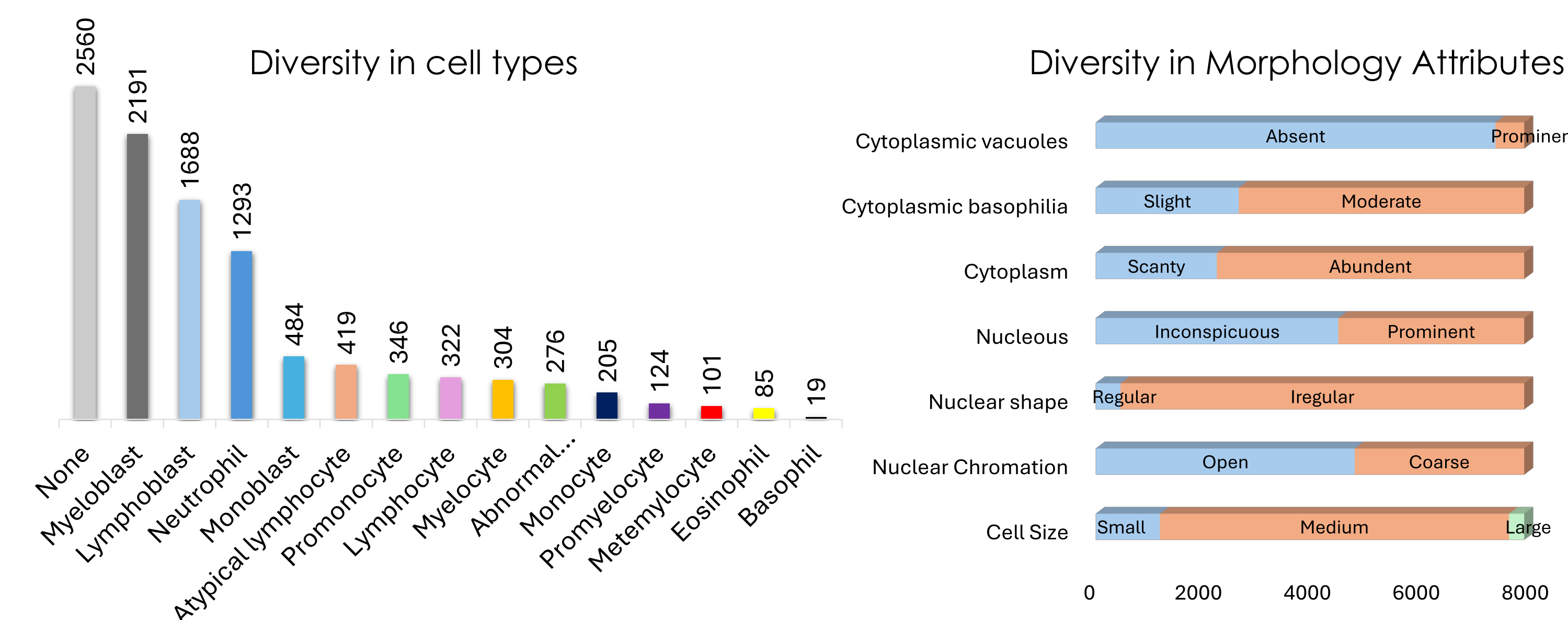
Monitoring Disease Progression

Evaluating Treatment Response

Assessing Bone Marrow Function

	WBC				
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

Distribution of WBC Types and Attributes

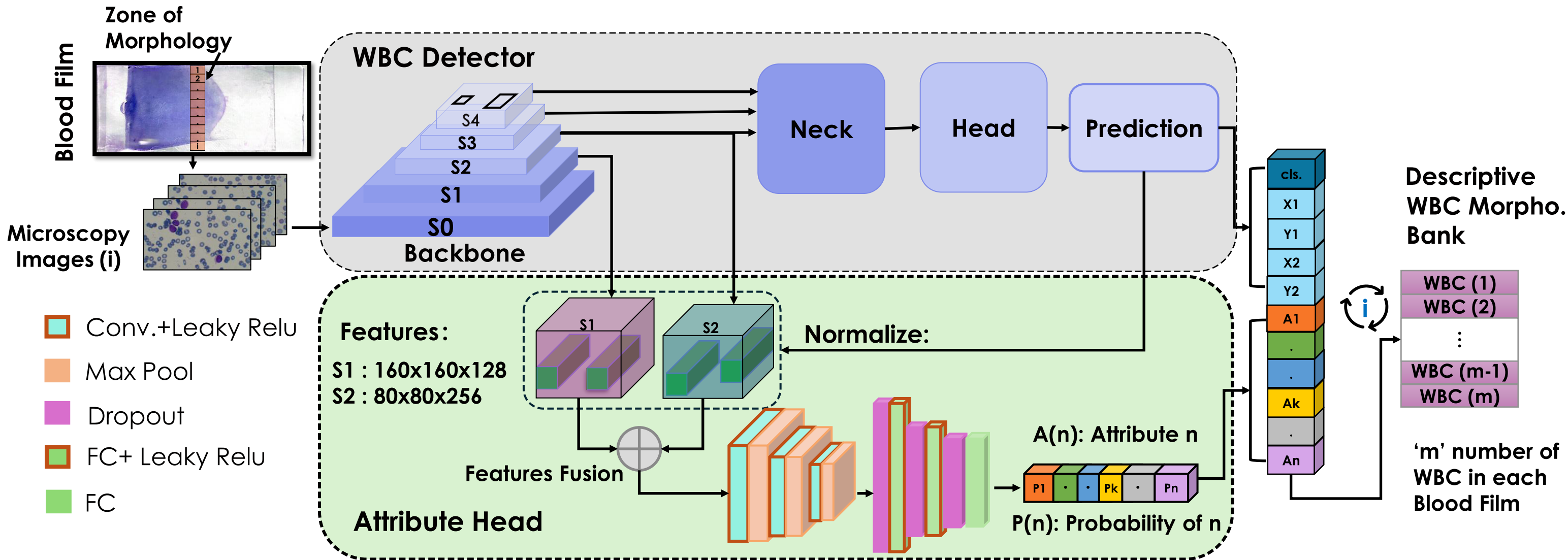


Comparison with Existing Datasets

Dataset	Type	Across Micro.	Multi. Cells in Image	BBX	Multi. Res.	No. of WBC's	Classes	Morphology
Ours (LeukemiaAttri)	Multi.	✓	✓	✓	✓	88.2 K	14	✓
IDB2 [10]	ALL	✗	✗	✗	✗	260	2	✗
LISC [17]	Normal	✗	✗	✗	✗	250	5	✗
Munich [15]	AML	✗	✗	✗	✗	18.3 K	15	✗
Raabin [9]	Normal	✗	✓	✓	✗	17.9 K	5	✗
HRLS [3]	Multi.	✗	✓	✗	✗	16 K	9	✗
WBCAt [25]	Normal	✗	✗	✗	✗	10.2 K	5	✓
IDB [20]	ALL	✗	✓	✓	✗	510	2	✗

LeukemiaAttri dataset consists of 28.8K images captured using low-cost and high-cost microscopes (2) at 3 different resolutions (10x, 40x, 100x) and 2 different cameras. (2.4K × 2 × 3 × 2)

Proposed Method (AttriDet)



Normalize
$$b_{norm}^{S_j,k} = \frac{x_k}{w_i} W_j, \frac{y_k}{h_i} H_j, \frac{w_k}{w_i} W_j, \frac{h_k}{h_i} H_j,$$

Similar size Features
$$F_{s,k} = ROI(S_j, b_{norm}^{S_j,k}) \quad \text{For } j \in \{1,2\}$$

Attribute loss
$$ASL = \begin{cases} \alpha = (i-p)^{r+} \log(p) & \text{if } p > 0 \\ \beta = (p\delta)^{r-} - \log(1-p\delta) & \text{if } p < 0 \end{cases}$$

Where
$$p\delta = \max(p - \delta, 0)$$

$$L = p\alpha - (i-p)\beta$$

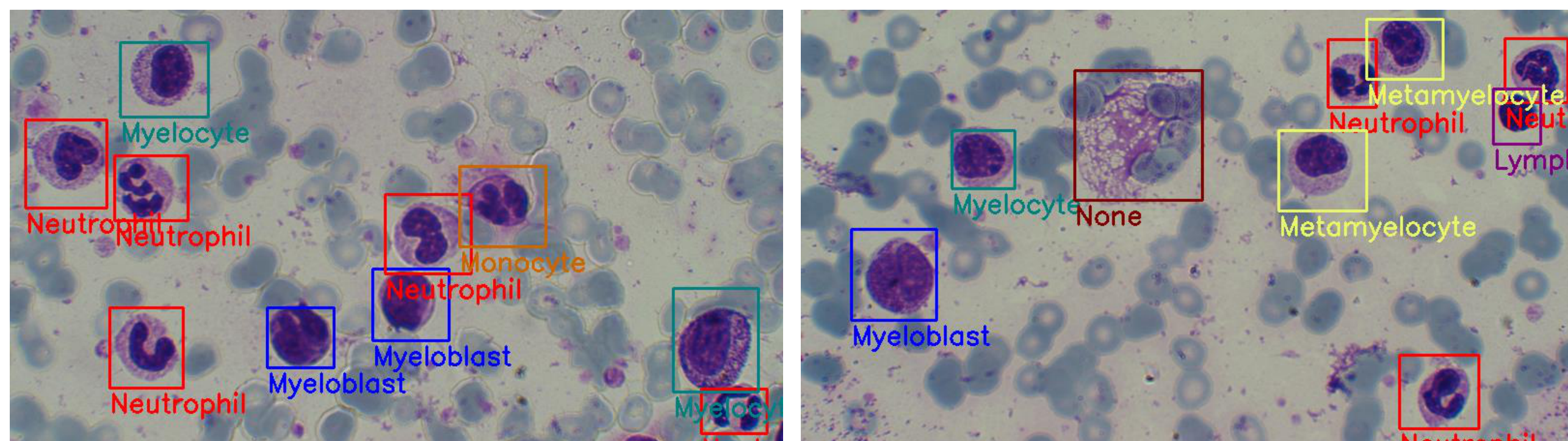
$$L_{mor} = \sum_{a=1}^A L(\sigma(p_a)m_a)$$

Total loss
$$L_{total} = L_{obj} + L_{cls} + L_{box} + L_{mor}$$

AttriDet Results

Method	NC	NS	N	C	CB	CV	WBC
CBM	21.9	96.2	41.8	77.2	70.2	3.33	27.6
AttriDet	73.9	95.9	54.3	89.7	83.6	29.1	28.2

Qualitative Results of AttriDet



AttriDet: Blood Film level Description

These blast cells are small to medium in size with scanty amount of cytoplasm, high N:C ratio, moderately condensed to dispersed chromatin nuclei and inconspicuous nucleoli. These cells contain vacuoles

WBC Detection Results on LeukemiaAttri

Method	Subset	mAP@50-95	mAP@50
Sparse R-CNN [23]	HCM_100x_C1	15.8	29.6
	HCM_100x_C2	21.3	36.7
	LCM_100x_C1	17.2	32.6
	LCM_100x_C2	14.5	25.9
FCOS [24]	HCM_100x_C1	16.4	31.8
	HCM_100x_C2	22.5	40.6
	LCM_100x_C1	17.5	33.9
	LCM_100x_C2	17.7	34.3
DINO [28]	HCM_100x_C1	17	33.8
	HCM_100x_C2	25.4	43.7
	LCM_100x_C1	17.5	34.3
	LCM_100x_C2	21.5	38.2
YOLOv5x [26]	HCM_100x_C1	20.9	38.8
	HCM_100x_C2	26.3	44.2
	LCM_100x_C1	20.7	39.5
	LCM_100x_C2	20.1	38.1

Domain Adoption results on LeukemiaAttri

Method	Train Subset	Test Subset	mAP@50-95	mAP@50
YOLOv5x (source only)	HCM_100x_C2	LCM_100x_C2	11	25.5
DACA	HCM_100x_C2	LCM_100x_C2	12.6	30.2
ConfMix	HCM_100x_C2	LCM_100x_C2	12.6	33.5

Conclusion and Future Directions

Dataset: Large-scale WBC Leukemia dataset with 12 subsets, using two microscopes and multiple cameras at 10x, 40x, and 100x resolutions.

Content: Contains 14 WBC types and 7 morphological attributes.

Method: AttriDet detects WBC types and their morphological features.

Benefit: Provides interpretable results to enhance doctors' confidence in AI-assisted diagnostics.

Impact: Supports research in explainable and robust Leukemia detection.

Reproducible code and Dataset set publicly available

