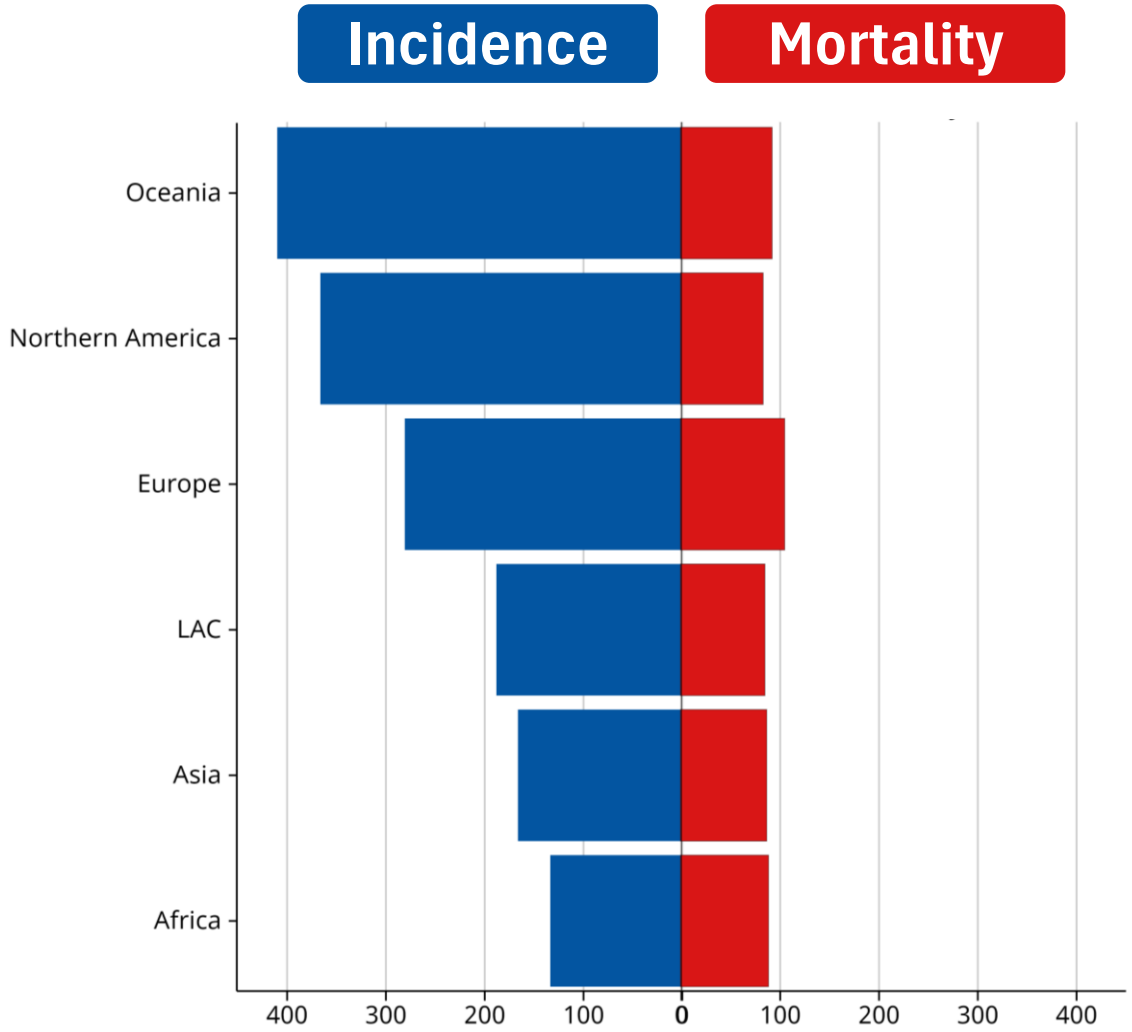
A Large-scale Multi Domain Leukemia Dataset for the White Blood Cells Detection with Morphological Attributes for Explainability

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

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

Leukemia

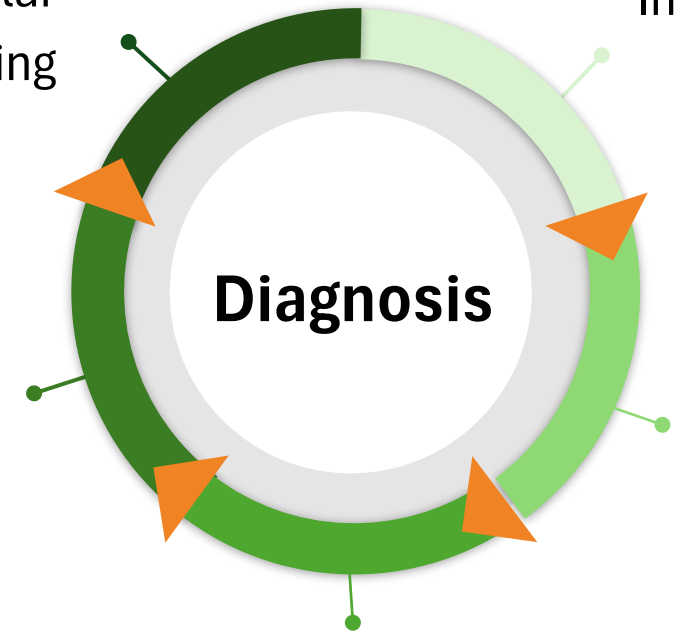
- 10^{th} Most Frequently Diagnosed Cancer
- Leading Cause of Cancer-related Deaths Under 40 Years
- Quick Divergence



Age-Standardized Rate (World) per 100 000, Incidence, Both sexes, in 2022

International Agency for Research on Cancer
World Health Organization

Problem Statement



Step 1:
Initial Evaluation

Step 2:
CBC and Peripheral Blood Smear Analysis

Step 3:
Flow Cytometry and Immunophenotyping:

Step 4:
Bone Marrow Aspiration and Biopsy

Step 5:
Cytogenetic and Molecular Testing

Problem Statement

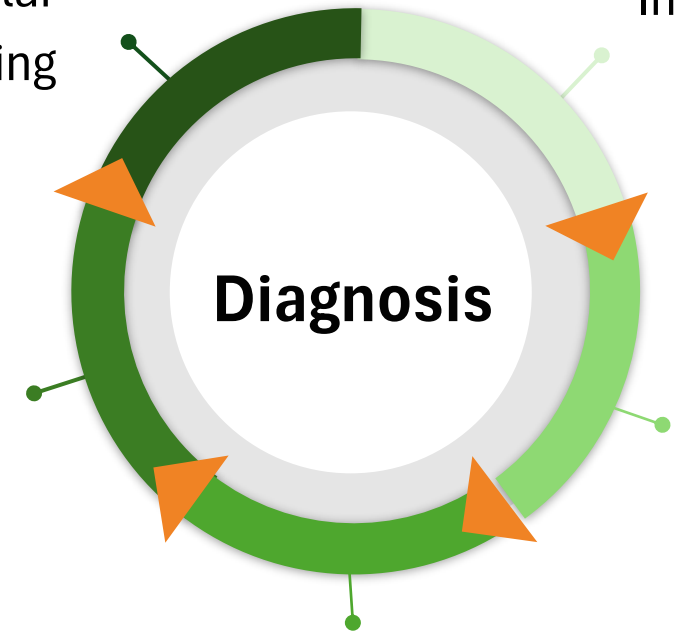
Step 1:
Initial Evaluation

Step 2:
CBC and **Peripheral Blood Smear Analysis**

Step 3:
Flow Cytometry and Immunophenotyping:

Step 4:
Bone Marrow Aspiration and Biopsy

Step 5:
Cytogenetic and Molecular Testing

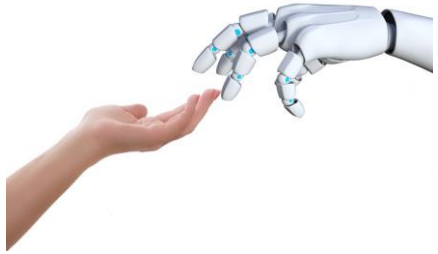


Motivation

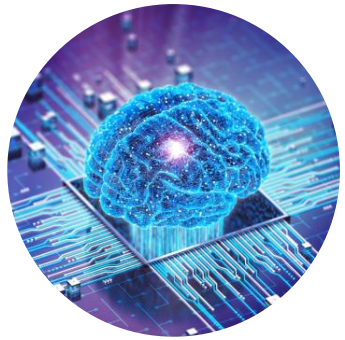
INFORMATION TECHNOLOGY UNIVERSITY

Microscopic Analysis

AI Assistance



Computer Aided Diagnosis (CAD)



- Increased Efficiency
- Enhanced Productivity
- Data Management
- Improved Accuracy

Knowledge Intensive

Shortage Medical Experts

Prone to Error

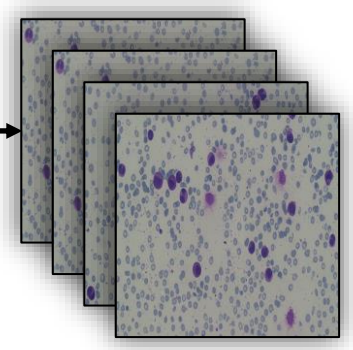
False Negative Rate
~ 3% to 15%

False Positive Rate
~ 10% to 28%

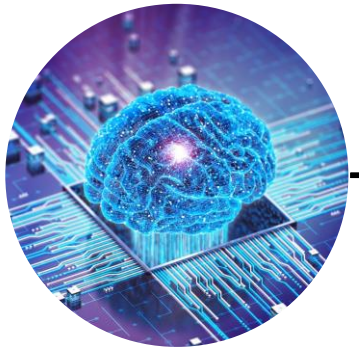
CAD System for Leukemia

INFORMATION TECHNOLOGY UNIVERSITY

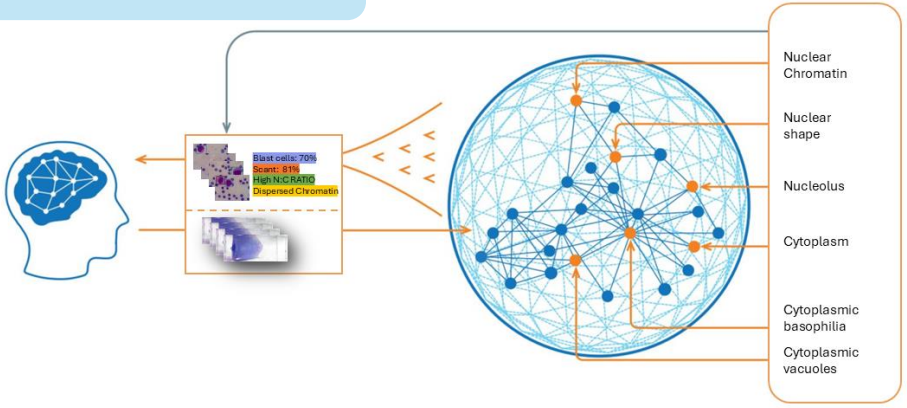
Large scale, Multi domain and **Broad Spectrum** Dataset



CAD System



Insightful Diagnosis



Comprehensive Device Adaptation

LeukemiaAttri:

Large Scale Multi Microscope, Multi Resolution, Multi Camera Dataset

Existing Leukemia / WBC Datasets

Dataset	Type	Across Micro.	Multi. Cells in image	BBX	Multi. Res.	No. of WBC's	Classes	Morphology
IDB [20]	ALL	✗	✓	✓	✗	510	2	✗
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]	ALL, AML, Normal	✗	✓	✗	✗	16 K	9	✗
WBCAtt [25]	Normal	✗	✗	✗	✗	10.2 K	5	✓

ALL: Acute Lymphocytic Leukemia
AML: Acute myeloid leukemia

- Single or Two types of leukemia
- Across the microscope (multimodality)
- Mostly the single cell in an image

- Missing the Localization of the WBC'c
- Single Resolution
- Limited number of WBC Classes

- WBC Morphology of abnormal cells

Existing Leukemia / WBC Datasets

Dataset	Type	Across Micro.	Multi. Cells in image	BBX	Multi. Res.	No. of WBC's	Classes	Morphology
Ours	Multi.	✓	✓	✓	✓	***88 K**	14	✓
IDB [20]	ALL	✗	✓	✓	✗	510	2	✗
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]	ALL, AML, Normal	✗	✓	✗	✗	16 K	9	✗
WBCAtt [25]	Normal	✗	✗	✗	✗	10.2 K	5	✓

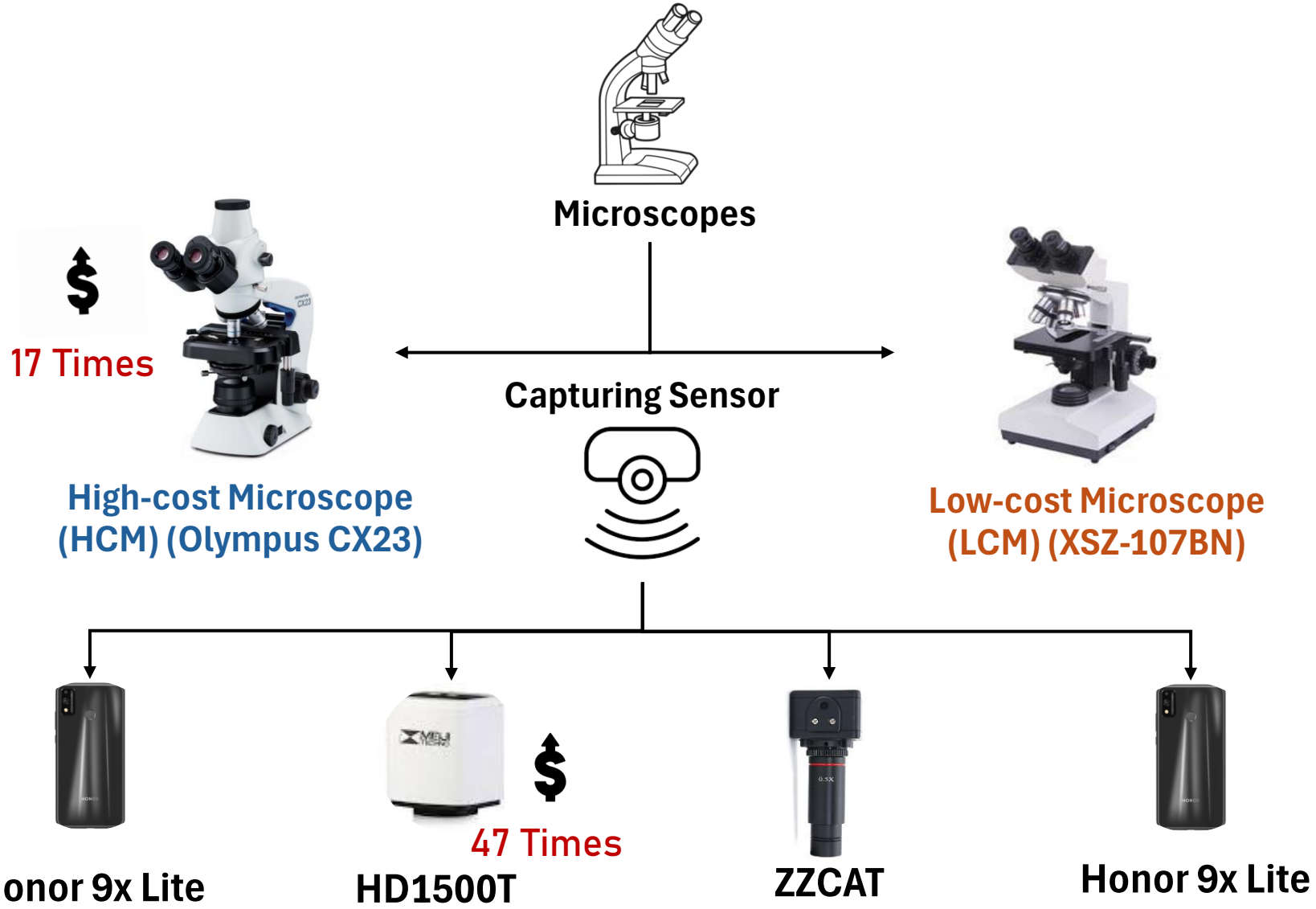
Multi. 1: ALL, 2: AML, 3: CML, 4: CLL, 5: APML

ALL: Acute Lymphocytic Leukemia
AML: Acute myeloid leukemia
CML: Chronic myeloid leukemia
CLL: Chronic lymphocytic leukemia
APML : Acute promyelocytic leukemia

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

LeukemiaAttri Collection

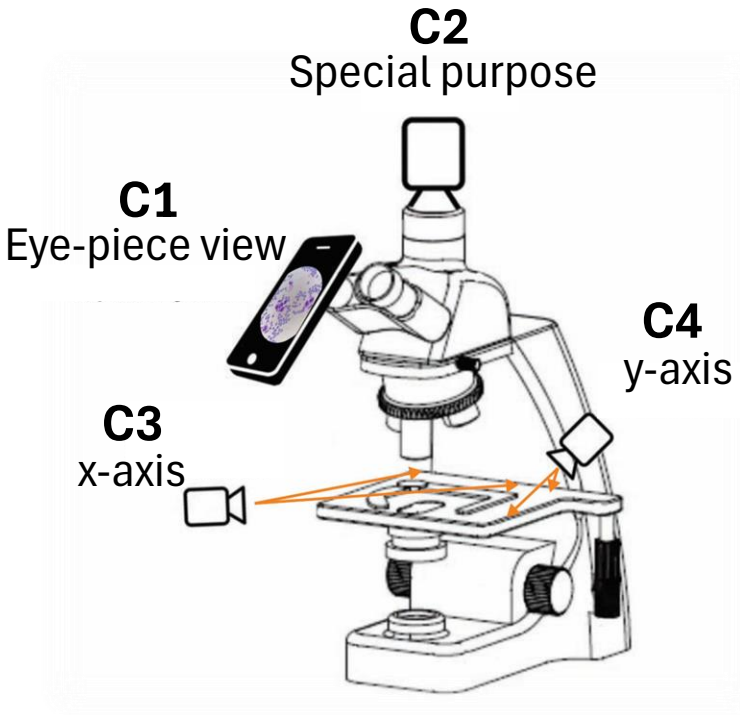
- Equipments



LeukemiaAttri Collection Procedure

Microscope: Olympus CX23
C1: Honor 9x Lite Mobile
C2: HD1500T
C3: Webcam1
C4: Webcam2

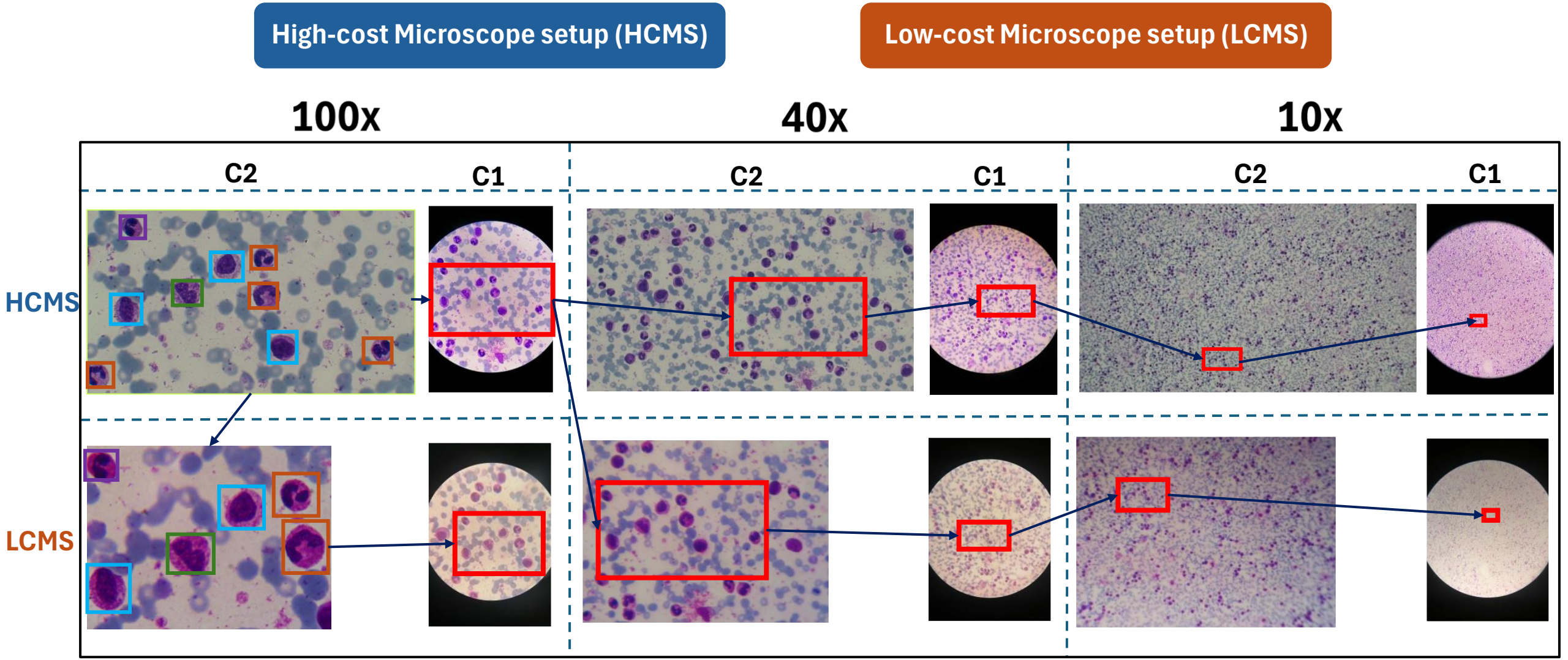
High-cost Microscope setup



Microscope: XSZ-107BN
C1: Honor 9x Lite Mobile
C2: ZZCAT 5MP
C3: Webcam1
C4: Webcam2

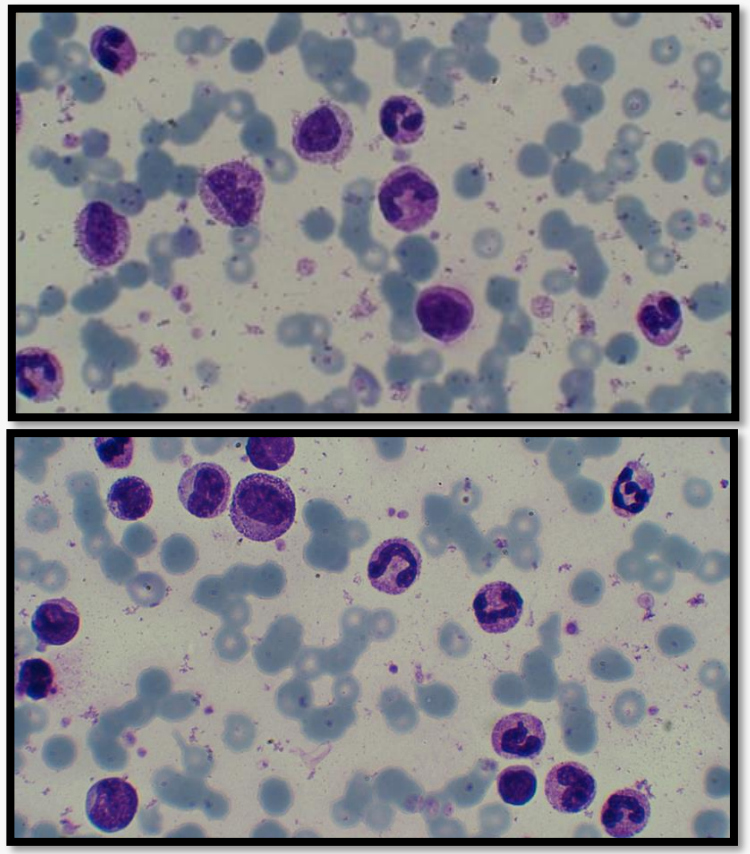
Low-cost Microscope setup

LeukemiaAttri Collection Procedure

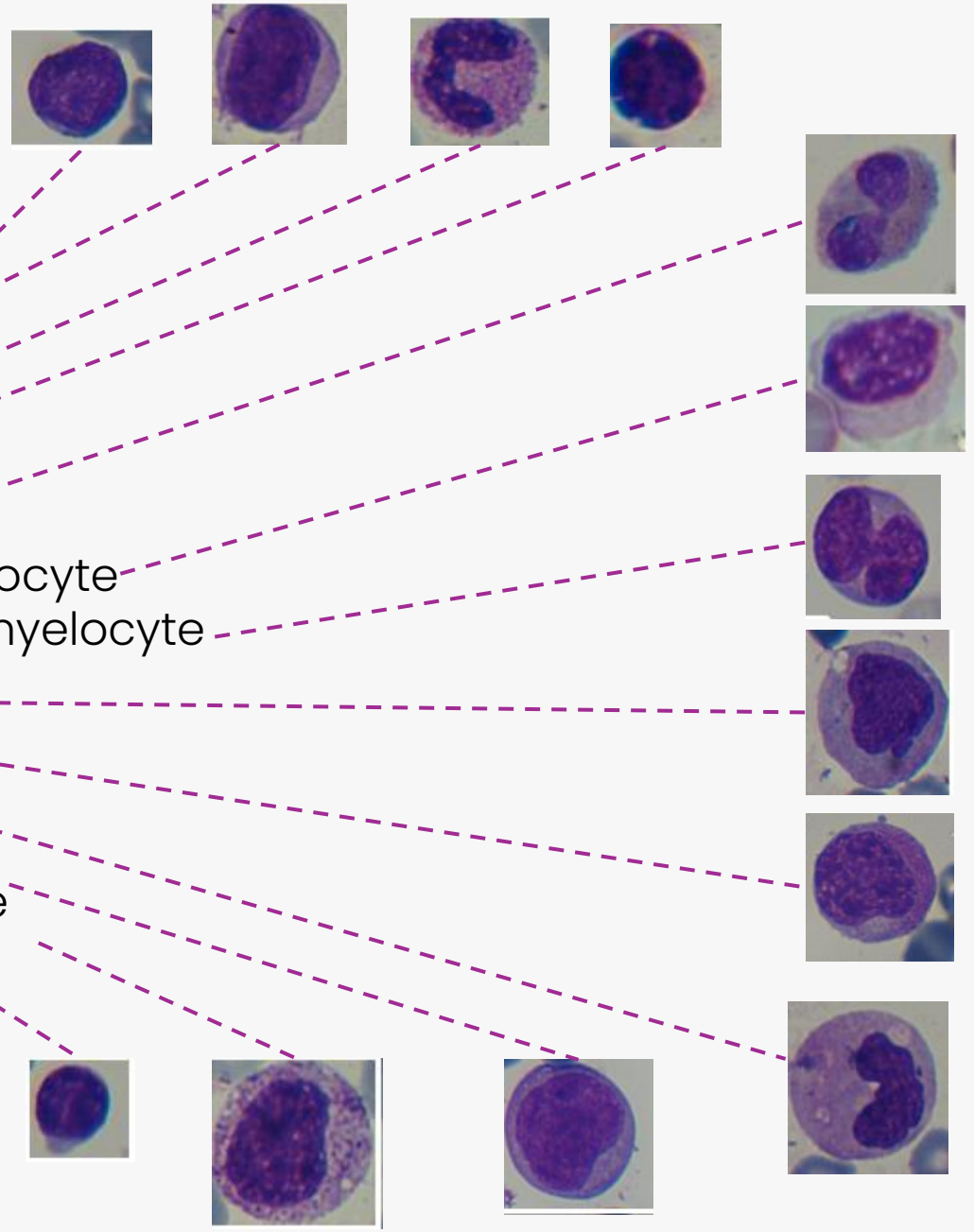


Note: *Hematologists annotate the highest quality images from the HCMS, and then we transfer them to other modalities.*

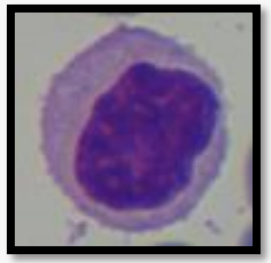
LeukemiaAttri Annotation



1. Lymphoblast
2. Myelobalst
3. Neutrophil
4. Basophil
5. Eosinophil
6. Atypical lymphocyte
7. Abnormal promyelocyte
8. Promonocyte
9. Myelocyte
10. Monocyte
11. Monoblast,
12. Metamyelocyte
13. Lymphocyte
14. None



LeukemiaAttri Morphology



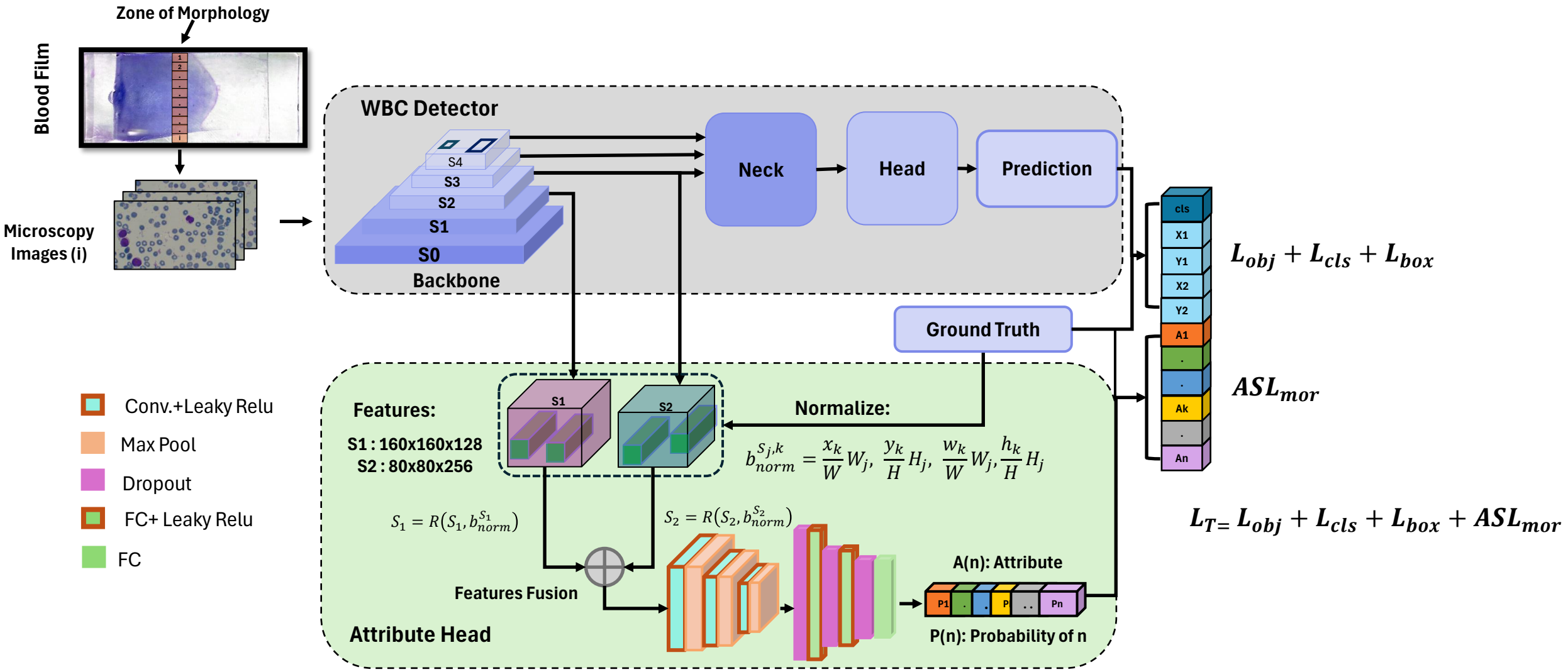
1. Lymphoblast
2. Myelobalst
3. Neutrophil
4. Basophil
5. Eosinophil
6. Atypical lymphocyte
7. Abnormal promyelocyte
8. Promonocyte
9. Myelocyte
10. Monocyte
11. Monoblast,
12. Metamyelocyte
13. Lymphocyte
14. None

I.	Cell size:	Medium to Large
II.	Nuclear chromatin:	Coarse
III.	Nuclear shape:	Irregular
IV.	Nucleolus:	Inconspicuous
V.	Cytoplasm:	Abundant
VI.	Cytoplasmic basophilia:	Moderate
VII.	Cytoplasmic vacuoles:	Absent

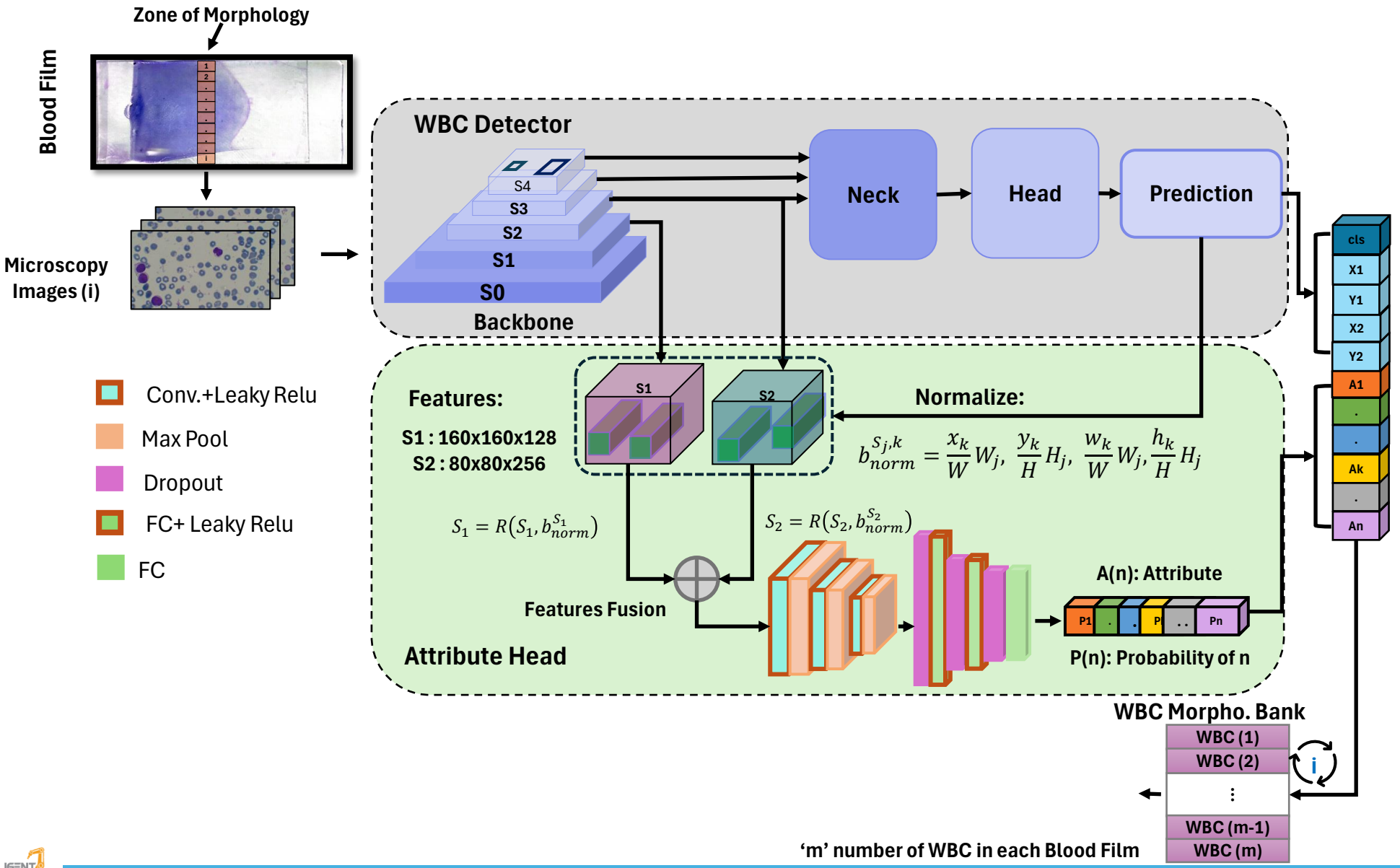
Proposed AttriDet

- Localize and classify WBC types
- Predict their morphology attributes

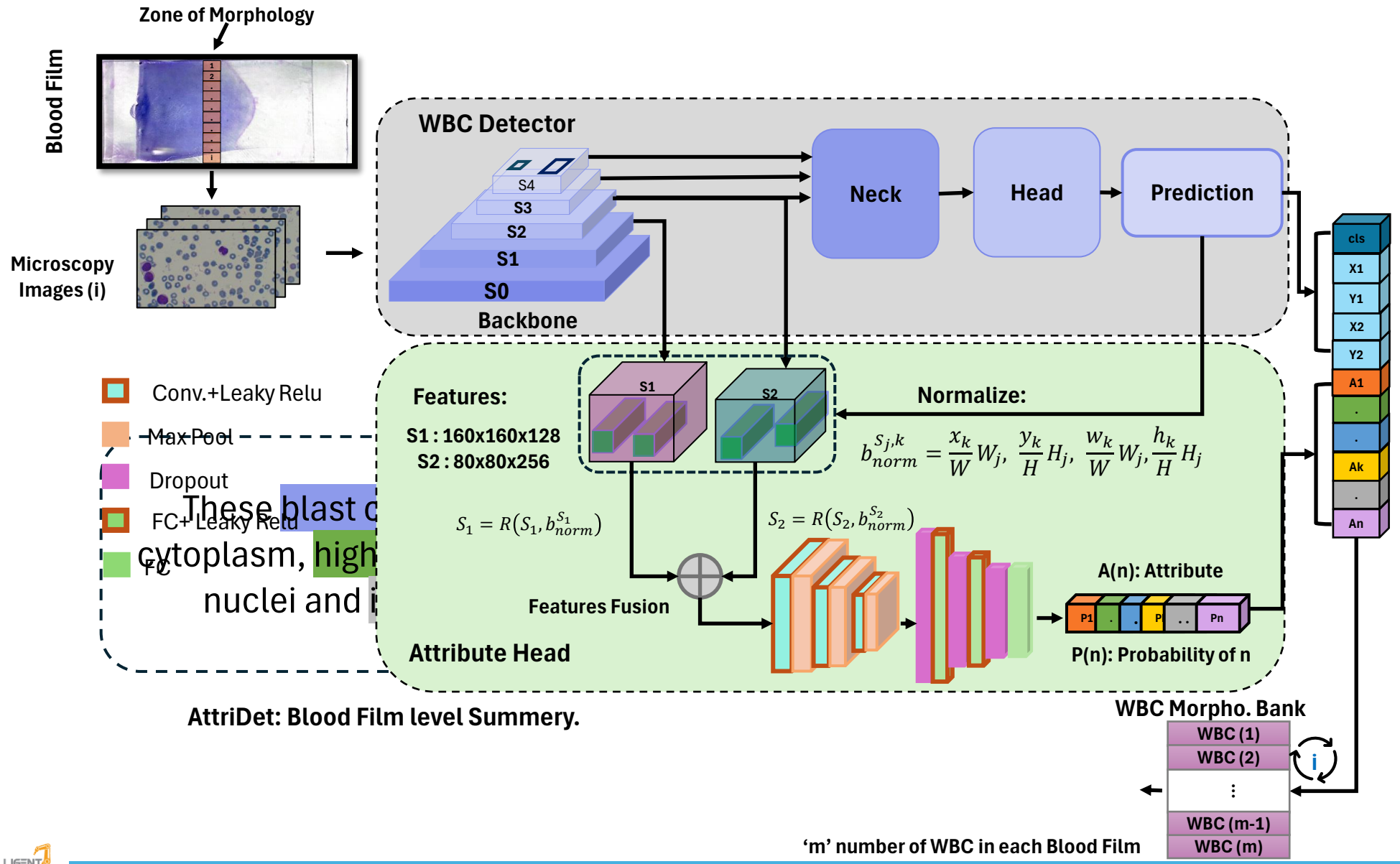
AttriDet : Proposed Model (Training)



AttriDet : Proposed Model (Inference)



AttriDet : Proposed Model (Inference)



Experimental Results

- WBC Detection
- Domain adaptation
- WBC Detection With Morphology

Experiments

WBC Detectors

Sparse R-CNN [23]	Two-stage process with a sparse proposal generation
FCOS [24]	Anchors free object detector
DINO [28]	Transformer-based architecture
YOLOv5x [26]	Single-stage architecture optimized for speed

WBC Detection Results on LeukemiaAttri

WBC Detectors Results

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

HCM: **Olympus CX23**
100x: Resolution
C2: Camera(HD1500T)
C1: Mobile

LCM: **XSZ-107BN**
100x: Resolution
C2: Camera(ZZCAT)
C1: Mobile

AttriDet Results on LeukemiaAttri

WBC Detection with Morphology

F1 Results of the morphology Attributes

Method	NC	NS	N	C	CB	CV	WBC mAP@50-95
YOLOv5x [26]	-	-	-	-	-	-	26.3

- **NC**: Nuclear Chromatin
- **NS**: Nuclear Shape
- **N**: Nucleus
- **C**: Cytoplasm
- **CB**: Cytoplasmic Basophilia
- **CV**: Cytoplasmic Vacuoles

AttriDet Results on LeukemiaAttri

WBC Detection with Morphology

F1 Results of the morphology Attributes

Method	NC	NS	N	C	CB	CV	WBC mAP@50-95
YOLOv5x [26]	-	-	-	-	-	-	26.3
CBM [8]	21.9	96.2	41.8	77.2	70.2	3.33	27.6

- **NC**: Nuclear Chromatin
- **NS**: Nuclear Shape
- **N**: Nucleus
- **C**: Cytoplasm
- **CB**: Cytoplasmic Basophilia
- **CV**: Cytoplasmic Vacuoles

- CBM **27.6**
- YOLOv5x 26.3

AttriDet Results on LeukemiaAttri

WBC Detection with Morphology

F1 Results of the morphology Attributes

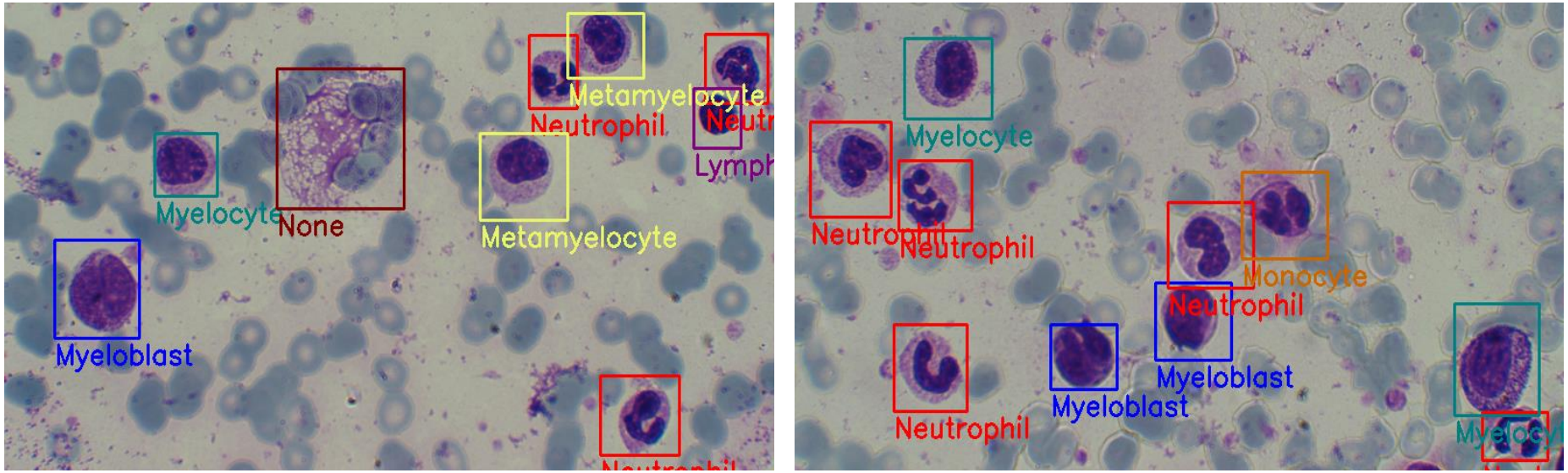
Method	NC	NS	N	C	CB	CV	WBC mAP@50-95
YOLOv5x [26]	-	-	-	-	-	-	26.3
CBM [8]	21.9	96.2	41.8	77.2	70.2	3.33	27.6
AttriDet	**73.9**	95.9	**54.3**	**89.7**	**83.6**	**29.1**	**28.2**

- **NC**: Nuclear Chromatin
- **NS**: Nuclear Shape
- **N**: Nucleus
- **C**: Cytoplasm
- **CB**: Cytoplasmic Basophilia
- **CV**: Cytoplasmic Vacuoles

- AttriDet (Ours) **28.2**
- YOLOv5x 26.3

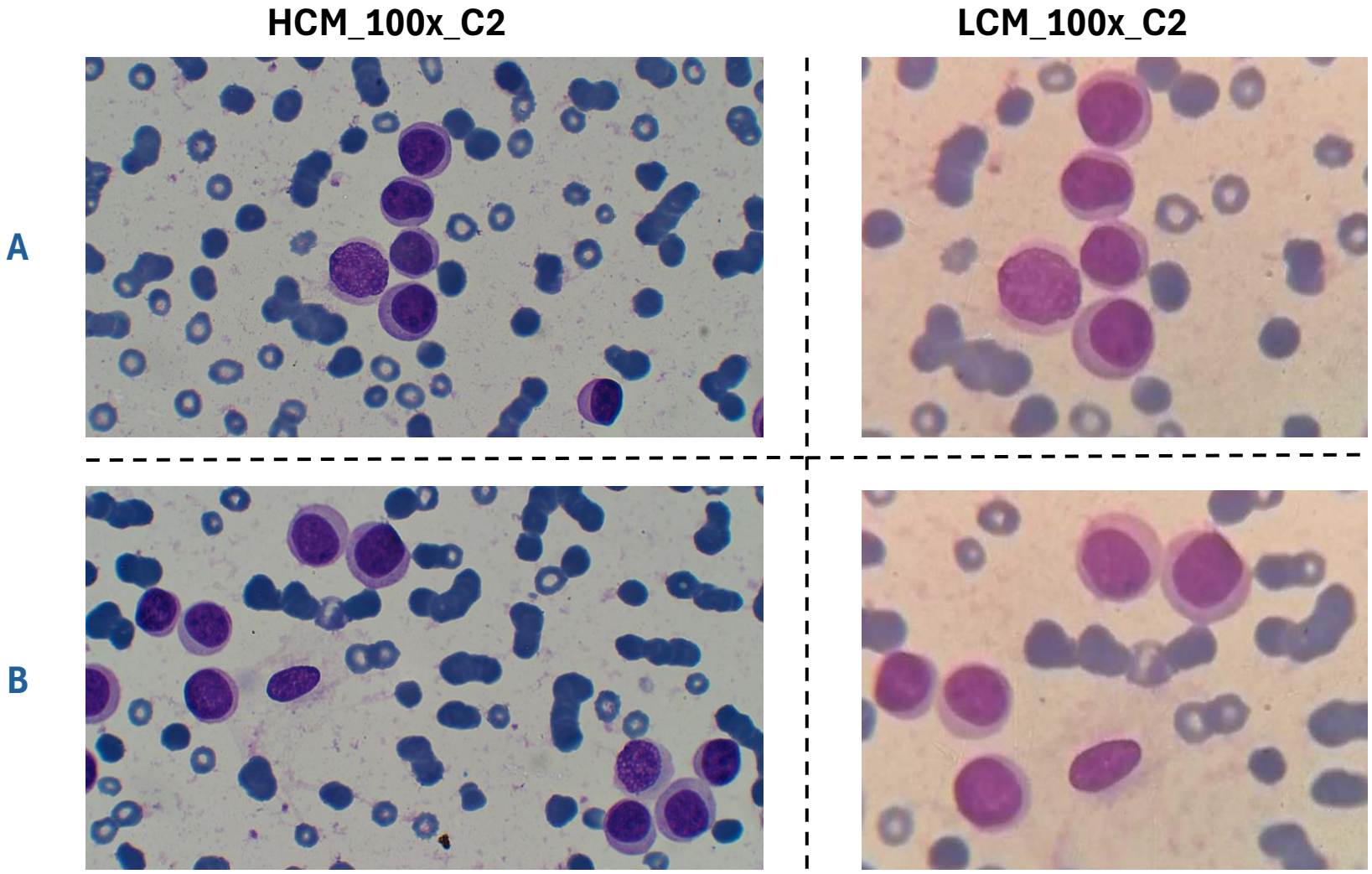
AttriDet Qualitative Results

AttriDet Detection



These blast cells (Myeloblast) are medium to large in size with abundant amount of cytoplasm, high N:C ratio, moderately condensed to dispersed chromatin nuclei and inconspicuous nucleoli. Some cells contain vacuoles

Domain Shift



- Different image resolution
- Different scaling
- Small Field of View
- Cells Enlarge in Size
- Low Quality

Domain Adaptation Base Lines

Source Data:

HCM: **Olympus CX23**
100x: Resolution
C2: Camera(HD1500T)

Target Data:

LCM: **XSZ-107BN**
100x: Resolution
C2: Camera(ZZCAT)

Method	Source Data	Target Data	mAP@50-95	mAP@50
YOLOv5x [26] (source only)	HCM_100x_C2	LCM_100x_C2	11	25.5
DACA [16]	HCM_100x_C2	LCM_100x_C2	12.6	30.2
ConfMix [23]	HCM_100x_C2	LCM_100x_C2	12.6	33.5

Releasing

Dataset:

- Utilizes two microscopes and multiple cameras
- Resolutions: 10x, 40x, and 100x
- **Content**:
  - Comprises 12 subsets
  - Contains 14 WBC types
  - Features 7 morphological attributes

WBC Detector with Attribute Predictor :

- AttriDet detects WBC types
- Identifies morphological attributes

- The code and Dataset Publicly available:

https://im.itu.edu.pk/medicalai-bloodcancerdataset-2024/

Contributions:

- **Dataset:**
  - Large-scale Multi Domain WBC Leukemia dataset
  - Utilizes two microscopes and multiple cameras
  - Includes resolutions of 10x, 40x, and 100x
  - Contains 14 WBC types
  - Features 7 morphological attributes
  - 28K images with 88k White blood cell
- **AttriDet:**
  - Localize and classify WBC types
  - Predict their morphology attributes

Thanks,
MICCAI Society for the Travel Grant

Any Questions?

Please Visit

Poster # W_AM_007

Day & Date: Wed. 9 oct 2024

Time: 10:00 to 11:30am