RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt2562

YOLO-Based Generalized Framework for Leukemia Cell Detection Using Unified Microscopic Image Datasets

Ratnamala Mantri (Paswan) [a, b, *], Rais Abdul Hamid Khan [a]

[a] SOCSE, Sandip University, Nashik, India

[b] Department of Computer Engineering, PICT, Pune, India

* Corresponding Author Email: ratnakumarsudhir@gmail.com

DOI: https://doi.org/10.54392/irjmt2562

Received: 02-06-2025; Revised: 11-09-2025; Accepted: 29-09-2025; Published: 15-10-2025

Abstract: Acute lymphoblastic leukemia is a kind of blood cancer that attacks the lymphoblast, a subgroup of white blood cells. Leukemia is a potentially lethal hematological cancer that requires prompt diagnosis. A skilled manual blood smear examination is one of the laborious and prone to human error conventional diagnostic methods. Although current automated methods developed by researchers use either single-cell or multi-cell pictures to detect leukemia cells, they frequently lack model generalization that perform better on heterogeneous datasets. They are also insufficient for deployment in real time. This study aims to develop generalized real-time system for detecting ALL cells from single and multi-cell microscopic blood smear images. The system utilizes three YOLO based state-of-the-art models: YOLO11, YOLOv8 and YOLOv5. The core novelty of this study lies in the creation of a unified dataset that integrates both single-cell and multi-cell microscopic blood smear images, this enables the model to learn generalized representations from diverse image contexts. Three datasets are merged to create the unified dataset, ALL-IDB1: multi- cell images, ALL-IDB2 & C-NMC-19: single-cell images. Image annotation and preprocessing are performed using Roboflow platform, while Google Colab is used for training and testing. These models are trained separately on individual datasets and the unified dataset.The performance of generalized YOLO models is assessed and contrasted against dataset-specific models using mAP@50 and recall metrics on the same set of unseen images from all three datasets.The experimental results indicate that generalized YOLOv8 model achieved notably high recall and competitive map@50, demonstrating strong adaptability and accuracy. These results highlight YOLOv8 as a promising solution for developing generalized model for leukemia cell detection.

Keywords: ALL IDB1, ALL IDB2, CNMC, Leukemia, Object detection, mAP@50, Medical image analysis, White blood cells, Real-time, Transfer learning, YOLO11s, YOLOv8, YOLOv5

1. Introduction

The rapid generation of aberrant white blood cells, or lymphocytes, causes leukemia, a type of cancer that can be seen in bone marrow and blood. These aberrant white blood cells hinder the production of red blood cells and platelets by the bone marrow and are unable to combat infection. Leukemia can be either acute or chronic, depending on the severity of the illness [1]. For better patient outcomes and prompt treatment, acute lymphoblastic leukemia (ALL), a potentially fatal hematological condition, has to be diagnosed early and accurately. The most frequent cancer to occur in children aged 0-14 is ALL, and estimations indicate that the number of cancer cases will increase by 12.8% in 2025 compared to 2020 [2].

It takes a lot of effort and time for skilled haematologists to manually examine blood smears under a microscope. Doctors' treatment decisions and the identification of medical diseases are directly impacted by how accurately this manual process is completed. White blood cell localization and classification are still difficult tasks, albeit [3]. Deep learning-based object detection models have drawn interest for medical imaging automation in order to get beyond these restrictions [4].

Object detection can be single stage detector or two stage detectors [5], here region will be proposed first and then it is classified [6]. A one-stage detector employs a direct learning framework to simultaneously classify and localize objects within an image. Compared to one-stage detector, two-stage detector offers faster processing speed [7].

YOLO family of designs has completely revolutionized object detection by introducing single stage detection mechanism [8]. YOLO family has undergone continuous improvement to enhance real-

time detection while maintaining high accuracy. YOLOv5 is recognized for its optimization and widespread adoption in real-time detection task. In contrast to anchor-based methods, YOLOv8 uses an anchor-free split Ultralytics head, which improves accuracy and streamlines the detection process. YOLO11, the latest YOLO model from Ultralytics [9], introduces refined architectural designs delivering faster processing speed and preserving the ideal ratio of performance to accuracy [10].

This study leverages three publicly available blood smear image dataset-ALL-IDB1, ALL-IDB2 [11], and C-NMC-19 [12]. We analysed these datasets for ALL cell detection using three YOLO based variants YOLOv5, YOLOv8 and YOLOv11.To check model generalization capability we also trained all three models on unified dataset: created by combining all three dataset into one dataset. Our intention is to demonstrate that the generalized model trained on unified dataset performs competitively or better on individual dataset.

1.1 Objective

1) To annotate and organize single-cell and multi-cell images from all three, ALL-IDB1, ALL-IDB2 and C-NMC-19 datasets and make the annotated data publicly available for research purpose. https://drive.google.com/drive/folders/1QgC_zY1NZX-fP0lis-Ti3CIBnLTPe_6C?usp=sharing
2) To develop a generalized ALL cell detection model capable of handling heterogeneous data by creating unified dataset with an equal number of annotated images from each of the three datasets. Train the each YOLO model-YOLOv5s, YOLOv8s and YOLO11s, on both individual dataset and the unified dataset.
3) To choose the best model in terms of generalization and detection accuracy by comparing the performance of generalized models to one another and to dataset-specific models on the same set of unknown datasets using common evaluation measures like recall and mAP@50.

2. Literature Survey

Recent advances in deep learning have significantly transformed medical image analysis, particularly in the detection and classification of diseases such as Leukemia, lung cancer, breast cancer and malaria. According to survey, the integration of deep learning approaches with transfer learning and customized CNN architectures has resulted in substantial improvements in diagnostic precision. Furthermore, hybrid approaches that combines CNNs with traditional machine learning algorithms have been effectively applied for automated detection of acute lymphoblastic leukemia cells and white blood cells. Transfer learning strategies, in particular, have consistently achieved high accuracy. However, one of the major challenges highlighted is the limited size of available medical image dataset, which can restrict the generalizability of the models [13].

Several studies have explored hybrid CNN approaches to improve ALL detection. Lamberti et al [14] proposed an explainable AI framework emphasizing interpretability for clinical use. Pradeep Das *et al.* [15] applied ResNet with OSL to enhance adaptability across dataset. AI-Bashir *et al.* [16] used CNN-based models to unified dataset-ALL-IDB2 and C-NMC-19 having single-cell images. Mantri *et al.* [3] combined K-means clustering with CNN for ALL multiclassification demonstrating the benefit of integrating tradition ML with deep learning. These methods lack cell localization and not optimized for real-time deployment.

Region -based CNN models have also been applied to blood cell analysis and ALL detection. Raina *et al.* (2020) presented a study that used the Faster R-CNN deep learning framework to detect blood cells. The model was created to enhance conventional detection methods by precisely identifying and categorizing blood cells in microscopic pictures. Faster R-CNN efficiently localized blood cells while preserving a high detection speed by utilizing the RPN. The study's exceptional precision and recall performance made it a viable method for automated haematological analysis. The approach will be enhanced in the future to better classify aberrant blood cells and to optimize it for real-time applications [17]. Using Mask R-CNN, Revanda *et al.* investigate a unique method for identifying and categorizing lymphoblastic leukemia cells (ALL) in white blood cell microscope pictures [18]. According to the findings, Mask R-CNN was able to correctly segment and categorize EVERY on the local dataset. The proposed model archives 83.72 % accuracy and 85.17 % precision. These region-based approaches provide precise detection and detailed cell-level analysis, though they generally require higher computational resources compared to single-stage detector.

YOLO architectures have emerged as prominent models for real-time detection and classification in medical image analysis. In order to diagnosis leukemia, Khandekar *et al.* Concentrate on an automated method that can identify and categorize blast lymphocyte cells from microscopic pictures. The single stage object detection model, YOLO4 (you only look once) was trained and evaluated on publicly available dataset, ALL-IDB1 and C-NMC-19 [19]. The use of object detection algorithms is investigated by Chen et al. Using a bounding box, a one-stage object recognition technique has been presented for the detection of leukemia cells in microscopic pictures.

Table1. Summary of existing approaches of ALL detection and classification

Author(s)	Method/Model Used	Application	Dataset	Key result	Limitation/Future Scope
Raina *et al*. (2020) [17]	Faster R-CNN with PRN	Blood cell Detection & classification	BCCD	mAP- 76.6%	Lower mAP and not optimized for real-time detection,
Khandekar *et al*. (2021) [19]	YOLO4	ALL detection.	C-NMC-19 ALL-IDB1	mAP-98.7% recall -96% mAP-96.06% recall -92%	Higher inference time restricts real-time applicability , lack of generalization capability as dataset-specific models and high computational cost (6000 epochs) risk of overfitting
Lamberti *et al.* (2022) [14]	Explainable AI	ALL detection	C-NMC-19 ALL-IDB2	Accuracy - 90.10% Accuracy - 100%	Unified approach was employed however its application restricted to single-cell image datasets, limiting its generalizability to multicell or heterogeneous dataset.
Pradeep Das *et al.* (2022) [15]	ResNet with OSL	ALL detection	ALL-IDB1 ALL_IDB2 C-NMC-19	Accuracy- 99.39% Accuracy - 98.21% Accuracy - 91.56	Dataset specific models. Lack of generalization and real-time capability.
Revenda *et al*. (2022) [14]	Mask R-CNN	ALL detection.	Local dataset of Blood smear images	Accuracy 83.72%	Lower accuracy, not scalable for real-time, limited generalization
Chen *et. al*. (2022) [20]	YOLOv5 YOLOv6 YOLOv7	ALL detection	ALL-IDB1	mAP YOLOv5- 90.15% YOLOv6- 64.4% YOLOv7- 58.3%	YOLOv5 outperformed YOLOv6 and YOLOv7. Inference time of YOLOv5- 12.3ms. Lack of Generalization.
Chou *et al*. (2024) [21]	YOLOv5 and YOLOv8	Esophageal cancer detection	Local dataset of esophageal images	YOLOv5 outperformed YOLOv8	Dataset specific models. The study did not report inference time.
Kundu *et al*. (2024) [22]	YOLOv8	Leukemia detection	Blood smear images	The research demonstrates the feasibility of YOLOv8 for real-time diagnosis of leukemia.	Lack of generalization.
Al-Bashir *et al.* (2024) [16]	CNN-based algorithms	ALL detection	ALL-IDB1	Accuracy- 94%	Unified approach was employed. however its application restricted to multi-cell image datasets, limiting its generalizability to single-cell or heterogeneous dataset
Mantri *et al.* (2025) [3]	K-Means , CNN	ALL Detection and Multiclassification	ALL-IDB1 and ALL-IDB2	Accuracy 100% L1-100% L2-99% L3-98%	High accuracy but risk of overfitting. lacks real-time and generalization capability.

YOLO5 is used for object detection on the ALL-IDB1-multi cell dataset. 97.2% multiclassification accuracy was attained [20].

Chou *et al.* (2024) evaluates the Spectrum-Aided Visual Enhancer integrated with YOLO-based deep learning frameworks YOLOv8 and YOLOv5 for esophageal cancer detection. The YOLO model, trained on enhanced images, achieves high precision and recall, outperforming conventional detection methods. The combination of spectral enhancement and deep learning significantly improves real-time cancer detection, reducing false positives and negatives [21]. Kundu *et al*. (2024) introduced a deep learning model using YOLOv8 to detect malignant patterns of microscopic stained blood images. By taking advantage of improved object detection properties of YOLOv8, the model recognizes white blood cells (WBCs) and differentiates between healthy and leukemic cells. Comprehensive evaluations showed that our approach can achieve high accuracy and robustness far exceeding conventional methods in both speed and accuracy. The research demonstrates the feasibility of YOLOv8 for real-time diagnosis of leukemia that helps to recognize the disease at the earliest stage and to make a clinical decision. In the future one could focus on further optimization of a subtype classification and to extend the data in order to generalize our findings across different clinical situations [22]. Table1 Summary of existing approaches for ALL detection and classification, highlighting methods, datasets, key results, and identified limitations.

The field of automated Acute lymphoblastic leukemia cell detection has witnessed significant progress due to the integration of deep learning and advanced object detection frameworks. However, a few persistent challenges remain unaddressed, particularly in building real-time, generalized, robust model using unified microscopic image datasets. Many studies primarily focus on training and evaluating models using isolated datasets such as ALL-IDB or CNMC. While these models always achieve high accuracy on a single dataset, they exhibit limited capacity to generalize across diverse datasets. Secondly, metrics such as accuracy is used for model evaluation but do not sufficiently consider the mAP score – a balanced indicator that consider both localization and classification accuracy. It has also been observed that YOLOv5 and YOLOv8 represent state-of-the-art models in medical image analysis, demonstrating strong potential for advancing leukemia detection task.

3. Proposed Methodology

To address the limitation of dataset specific models, the proposed study introduces a YOLO -based generalized framework shown in figure 1, for real time detection of leukemia cell by leveraging a unified set of microscopic image dataset. You Only Look Once (YOLO), is single stage neural network-based architecture.

YOLO is used for object localization, as well predicts the class probabilities of the object. Since the first model of YOLO i.e. YOLOv1 [23] has released in 2015, multiple additional variants of the same model have been developed by different groups with distinct goals, each building on and improving the previous one [8-10].

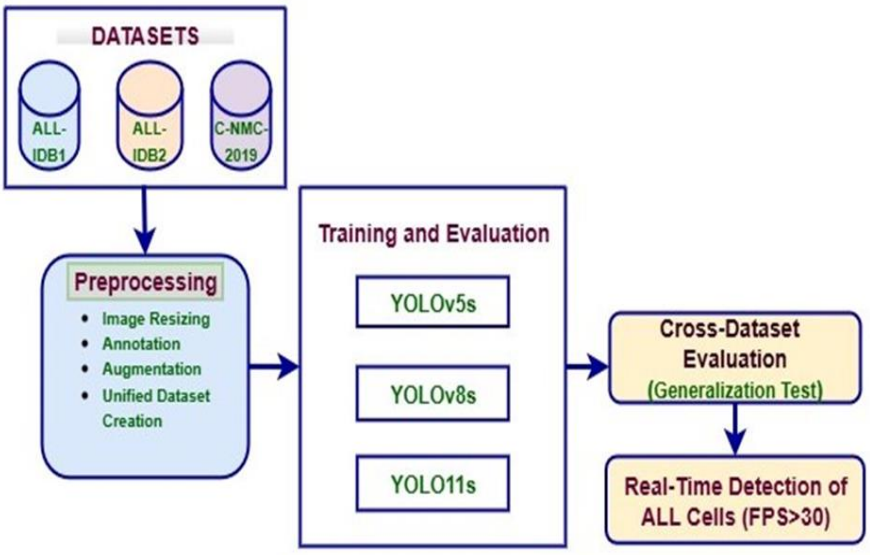

Figure 1. Proposed generalized architecture for real-time detection of leukemia cells

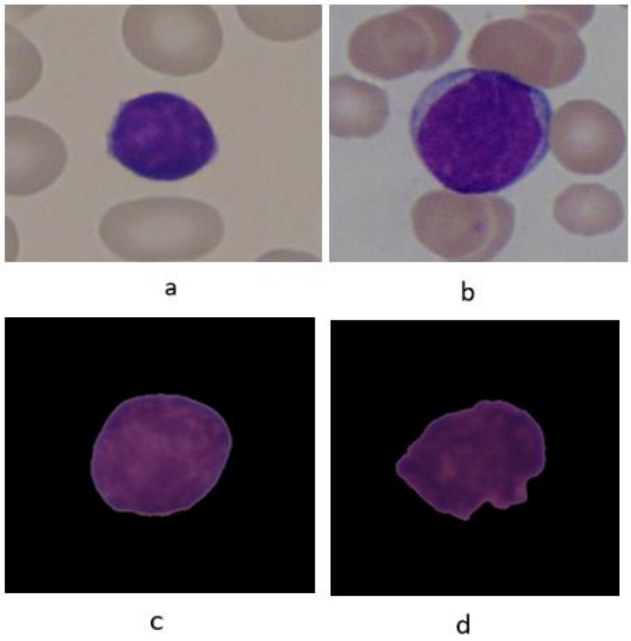

Figure 2. Representative images of healthy cells (a, c) and Acute Lymphoblastic leukemia (ALL) cells (b, d) from ALL-IDB2 & C-NMC-19 dataset

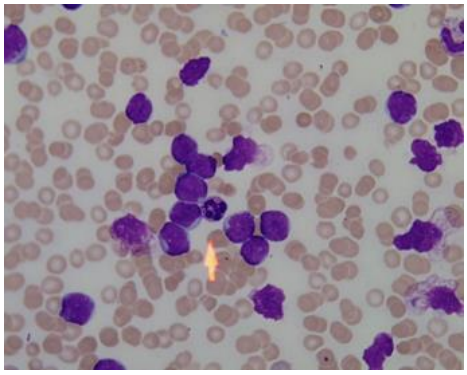

Figure 3. Sample multi-cell peripheral blood smear image from ALL-IDB1 dataset

3.1 Dataset

We made use of the ALL-IDB1 & ALL-IDB2 databases, which are publicly accessible and contain microscopic pictures of blood samples from both normal and ALL patients [11]. Experts in oncology provide the classification or location of ALL lymphoblasts for every image in the dataset. The 108 photos that make up ALL-IDB1 each contain several healthy and lymphoblast cells, with experts marking the blasted cells. The 260 photos in ALL-IDB2 are a compilation of cropped normal and blast cell areas of interest from the ALL-IDB1 dataset. The IEEE ISBI hosted the C-NMC 2019: medical imaging competition [12] of the 118 participants in the dataset, 69 were diagnosed with ALL and 49 with Hem. Sample blood smear images of leukemia patients and healthy people are shown in Figure 2. While photos (c) and (d) are from the C-NMC-19 dataset, images (a) and (b) are from the ALL-IDB2 dataset. A sample from the ALL-IDB1 dataset, which includes 108 multi-cell microscopic blood smear images, is presented in Figure 3.

3.2 Image Annotation, Augmentation, unified dataset preparation

Image annotation is a critical step in building the custom dataset for YOLO model training since it directly affects the accuracy and performance of the object recognition model. The model will be able to locate objects in the image with precision if the annotation is accurate. We use two classes to annotate microscopic images from all datasets: class 0-ALL (leukemia cell) and class 1-Hem (healthy cell). Each image in the YOLO annotation format has to have a corresponding.txt file. The annotation data-object class, height, width, and the X and Y coordinates of the bounding box center are all contained in this.txt file. Figure 4 displays an example of an annotated image file. Domain experts then examined and validated the annotated data to make sure that the class labels and bounding box placement were correct and consistent.

Since ALL-IDB1 and ALL-IDB2 datasets are a small dataset consisting of 108 and 260 images respectively, we augmented the images to make up.

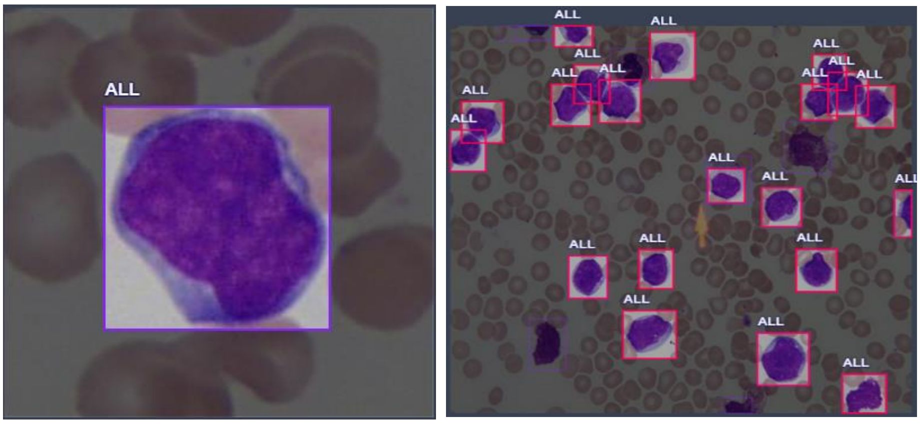

Figure 4. Example of YOLO Annotation format for microscopic blood smear image

We applied different augmentation techniques like horizontal & vertical flipping then rotated clockwise, counterclockwise by 90°. After data augmentation, the final dataset consists of 257 images in ALL_IDB1 and 584 images in ALL-IDB2 dataset. In this work, Roboflow is used for annotation and augmentation purposes. Roboflow framework can be used for dataset management like annotating, preprocessing, and augmentation. To address the limitation of data-specific models, this research consolidates multiple publicly available microscopic image datasets namely, ALL_IDB1, ALL-IDB2, and C-NMC-19 into unified dataset. The annotation format of all datasets was standardized to YOLO format to ensure consistency. To maintain the class balance across the unified dataset, an equal no of images, approx. 200 per class were selected from each dataset. Since ALL-IDB1 contains fewer images, the uniform sampling helped create a balanced dataset. A corresponding YAML configuration file was generated to reflect unified data structure.

3.3 YOLO architecture

You Only Look Once (YOLO) suggests utilizing an end-to-end neural network to forecast bounding boxes and class probabilities simultaneously. It varies from prior object detection methods, which used two stage detector classifiers for detection. Since the initial release of YOLO in 2015 [23], multiple additional variants of the same model have been developed by different groups with distinct goals, each building on and improving the previous one. YOLO11s is the latest release of YOLO series [8].

The architectural progression from YOLOv5 to YOLO11, demonstrates the significant improvement in feature extraction, refinement, and object detection. All three models YOLOv5, YOLOv8, and YOLO11s share a common high-level architecture composed of three core components-Backbone, Neck and Head [9].

In YOLOv5, the backbone utilizes a CSPDarknet53for initial feature extraction. The Neck incorporates a PANet-path aggregation network for path-wise feature aggregation across scale. The Head-detection layer responsible for bounding box regression and classification using anchor-based predictions.

YOLOv8 introduces architectural advancements to improve speed and accuracy. Its backbone is lightweight and optimized version of CSPDarknet. The Neck adopts FPN-feature pyramid network, improving multiscale feature fusion. The Head is decoupled into separate branches for classification and localization, the model is made anchor free, enabling faster training convergence and simple label assignment strategies.

YOLO11, shown in figure 5 introduces key innovations tailored for enhanced spatial understanding and computational efficiency. The backbone employs convolution layer integrated with C3k2 block. For spatial attention of feature, YOLO11s introduces C2SPA (cross stage partial with spatial attention) block in addition to existing SPPF-Spatial Pyramid Pooling-Fast block. Neck block focuses on efficient and faster way to focus on spatial feature using C3k2 block and distinguished C2SPA.The last component Head is responsible for anchor-based predictions, predicting bounding box and class probabilities. Further CBS layer is added for feature selection. This layer is responsible for passing feature map to subsequent bounding box prediction and classification layer [10].

3.4 Evaluation metrics

Mean average precision (mAP), which gauges accuracy and localization (IoU) for each category—in our case, ALL (leukemia) and Hem (healthy)—is a metric frequently applied to assess object detection models. The performance of several models is compared using mAP as a parameter.

The intersection over union (IoU), defined in Eq. 1 as the ratio of the intersection area of the predicted bounding box and real bounding box to their union, is used to define positive prediction using average precision metrics at different thresholds, often between 0.5 and 0.95. The AP@50 statistic, which is commonly used to evaluate the model's accuracy, is defined in Eq.

2. N is the number of classes, and AP50 is the average accuracy value at IoU=0.5 for each class [8, 9].

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (1)$$

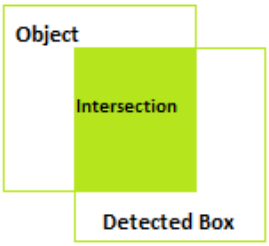

Figure 6. Intersection over Union

$$mAP@50 = \frac{1}{N}\sum_{i=0}^{N} AP_{50,i} \quad (2)$$

mAP uses bounding box loss to enhance the localization and cross entropy loss to improve classification accuracy.

Recall is the metric which is influenced by the Objectness loss and localization loss (IoU). Objectness loss tells the model whether an object is present in a bounding box or not and localization loss helps to accurately localize the bounding box.

4. Experimental Results and Analysis

This study employs YOLO models-YOLO11s, YOLOv8s, and YOLOv5s [24-26] (small variant) for leukemia cell detection. These models were trained on three datasets, ALL-IDB1, ALL-IDB2, and C-NMC-19 as well as unified dataset combining all three. Dividing the total dataset images into a training, validation, and testing sets as summarized in Table 2.

The model is trained using Google Colab Pro, which includes 12.7 GB of system RAM, 15.0 GB of GPU RAM, and 235.7 GB of hard drive space. The model is fine-tuned using 50 epochs, a batch size of 16 samples, and an input size of 640 pixels. Figure 7-9 displays the training and validation loss curves as well as the precision and recall trends for the YOLOv5s, YOLOv8s, and YOLO11s models trained on the Unified dataset.

Table 2. Dataset Split

Dataset	Training set		Validation		Testing		Total
	ALL	HEM	ALL	HEM	ALL	HEM	
ALL-IDB1	7	225	21	20	10	11	257
ALL-IDB2	240	257	24	4	0	19	584
C-NMC-19	199	200	20	0	0	20	479
Unified dataset	519	701	66	5	9	56	1466

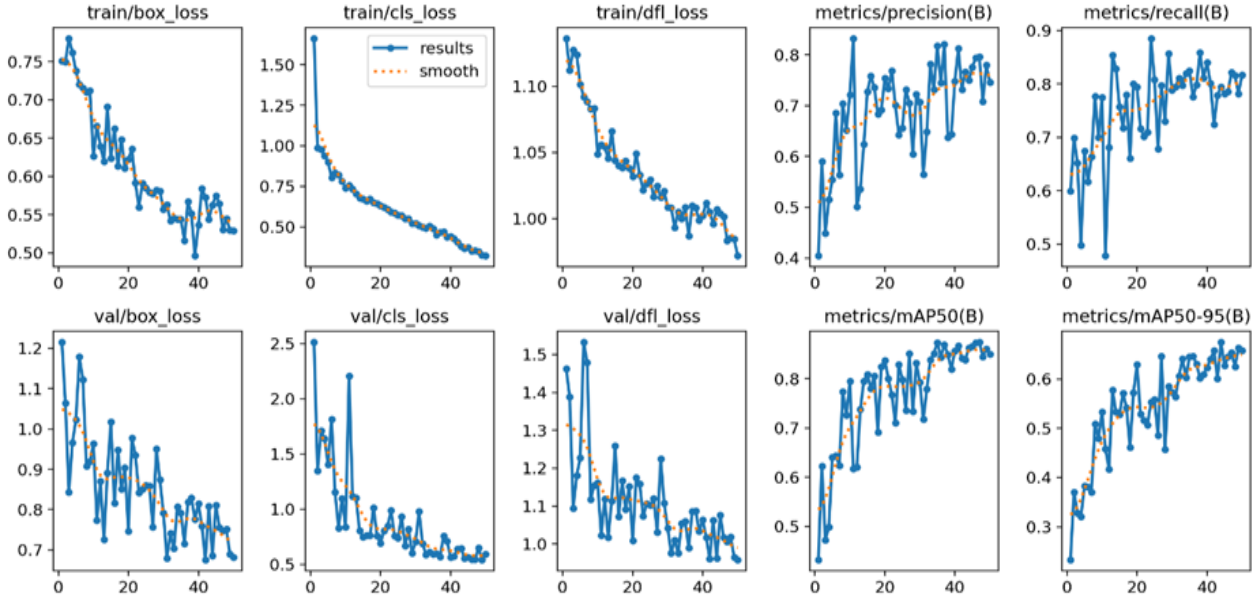

Figure 7. YOLOv5s: Training and validation loss curves, precision, recall and mAP@50 Vs across epochs on unified dataset

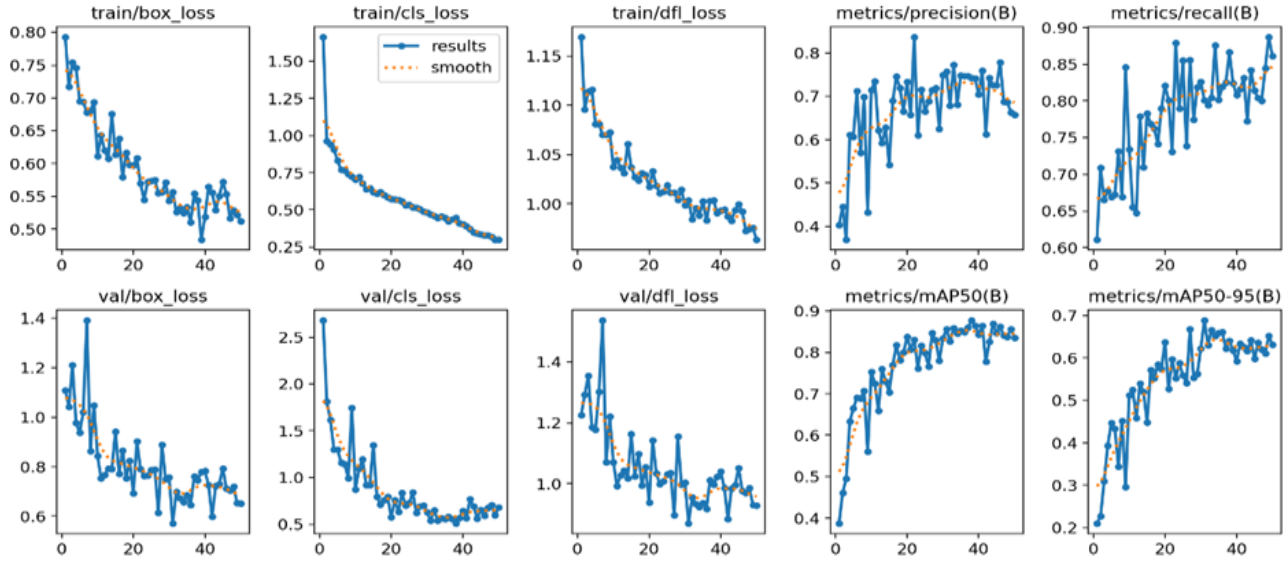

Figure 8. YOLOv8s: Training and validation loss curves, precision, recall and mAP@50 Vs across epochs on unified dataset

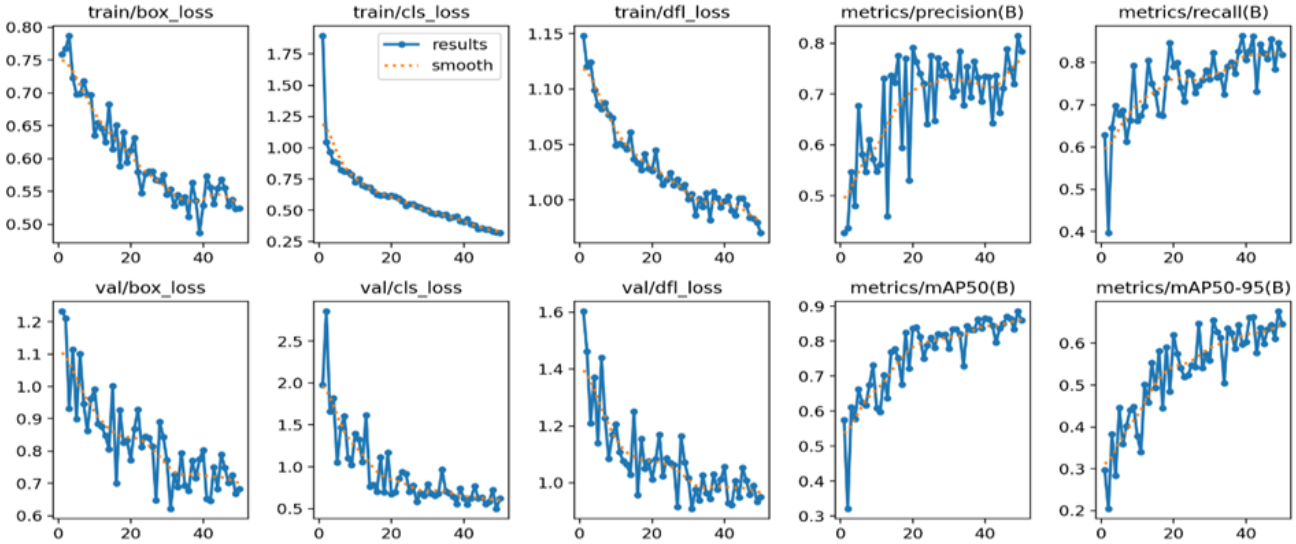

Figure 9. YOLO11s: Training and validation loss curves, precision, recall and mAP@50 Vs across epochs on unified dataset

Table 3. Performance comparison of YOLO11s, YOLOv8s and, YOLOv5s models on ALL-IDB2 dataset using Recall and mAP@50

Class	Recall			mAP@50		
	YOLO11s	YOLOv8s	YOLOv5s	YOLO11s	YOLOv8s	YOLOv5s
ALL	100	100	100	98.1	98.8	98.8
Hem	78.9	94.7	93.9	96.6	98.8	98.8
Total	89.5	97.4	97	97.3	98.8	98.8

Table 4. Performance comparison of YOLO11s, YOLOv8s and, YOLOv5s models on C-NMC-19 dataset using Recall and mAP@50

class	Recall			mAP@50		
	YOLO11s	YOLOv8s	YOLOv5s	YOLO11s	YOLOv8s	YOLOv5s
ALL	100	80	75	90.7	93.6	89.2
Hem	76.6	100	100	95.9	96.5	92.1
total	88.3	90	87.5	93.3	95	90.6

Table 5. Performance comparison of YOLO11s, YOLOv8s and, YOLOv5s models on ALL-IDB1 dataset using Recall and mAP@50

Class	Recall			mAP@50		
	YOLO11s	YOLOv8S	YOLOv5s	YOLO11s	YOLOv8s	YOLOv5s
ALL	100	89.7	100	97.4	95.3	93.5
Hem	66.7	72.2	72.2	71.3	74.2	69.4
total	83.3	81	86.1	84.3	84.8	81.4

Table 6. Performance comparison of YOLO11s, YOLOv8s and, YOLOv5s models Trained and tested on unified dataset using Recall and mAP@50

Class	Recall			mAP@50		
	YOLO11s	YOLOv8s	YOLOv5s	YOLO11s	YOLOv8s	YOLOv5s
ALL	93.8	91.3	79.9	95.2	95.9	94.6
Hem	79.2	80.2	77.4	71.5	76.7	68.8
total	86.5	85.7	78.6	83.3	86.3	81.6

Table 7. Evaluation and comparison of generalized YOLO models versus dataset specific YOLO models on individual datasets using mAP@50

Dataset/ Model	YOLO11s (Generalized)	YOLOv8s Unified	YOLOv5 Unified	YOLO11s	YOLOv8s	YOLOv5s
ALL-IDB1	77.4	86.9	76.2	84.3	84.8	81.4
ALL-IDB2	97.8	98.1	97.6	97.3	98.8	98.8
CNMC_2019	90.3	91.1	88.6	93.3	95	90.6

Table 8. Evaluation and comparison of generalized YOLO models versus dataset specific YOLO models on individual datasets using Recall

Dataset/ Model	YOLO11s Unified	YOLOv8s Unified_	YOLOv5 Unified_	YOLO11s	YOLOv8s	YOLOv5s
ALL-IDB1	90.9	93.1	76.2	83.3	84.8	81.4
ALL-IDB2	96.2	97.2	91.9	89.5	97.4	97
CNMC_2019	88.8	94.3	90.6	88.3	90	87.5

The comparative performance of YOLOv5s, YOLOv8s and YOLO11s was analysed across ALL-IDB1, ALL-IDB2, C-NMC-19 and the unified dataset using recall and mAP@50(Table 3-6). These metrics were chosen because mAP@50 is widely accepted for evaluating object detection models, providing good trade-off between localization and classification accuracy, especially in small-object detection scenarios like leukemia cell detection.

YOLOv8s consistently achieves the highest or near-highest score across datasets, particularly excelling on ALL-IDB2 with an overall recall of 97.4% and mAP@50 of 98.8%. on C-NMC-19, it attains the highest mAP@50 (95%) and perfect recall for the Hem class, demonstrating robustness in detecting minority classes. Similarly, on the unified dataset, YOLOv8s records an mAP@50 of 86.3%, outperforming YOLOv5s and YOLO11s, confirming its superior generalization ability.

The improved performance of YOLOv8s can be attributed several architectural enhancement: anchor-free detection head, decoupled detection head, dynamic label assignments and advanced data augmentation which reduce overfitting and enhanced robustness on heterogeneous data. However, its relative performance dips on multi-cell images i.e. ALL_IDB1, because densely packed and overlapping cells pose challenges for anchor-free center prediction [8, 9].

On the contrary, YOLO11s introduces more powerful components like C2f and C2SPA blocks to promote spatial attentions, which is beneficial to feature representation for complex structures. Yet its recall on the Hem class is consistently lower than YOLOv5s and YOLOv8s across datasets (e.g., 78.9% on ALL-IDB2 and 66.7% on ALL-IDB1). This is probably due to the sensitivity of the technique to the class imbalance and the increased architecture complexity that can drive the detection towards the predominant classes (i.e., ALL) [9, 10].

Table 9. Inference speed and Real time performance of YOLO models, trained on Unified dataset

Model	Preprocess(ms)	Inference(ms)	Postprocess(ms)	Frames Per Second (FPS)
YOLOv5s	1.0	4.7	9.5	212.77
YOLOv8s	0.5	5.8	7.2	172.41
YOLO11s	0.3	6.3	6.2	158.73

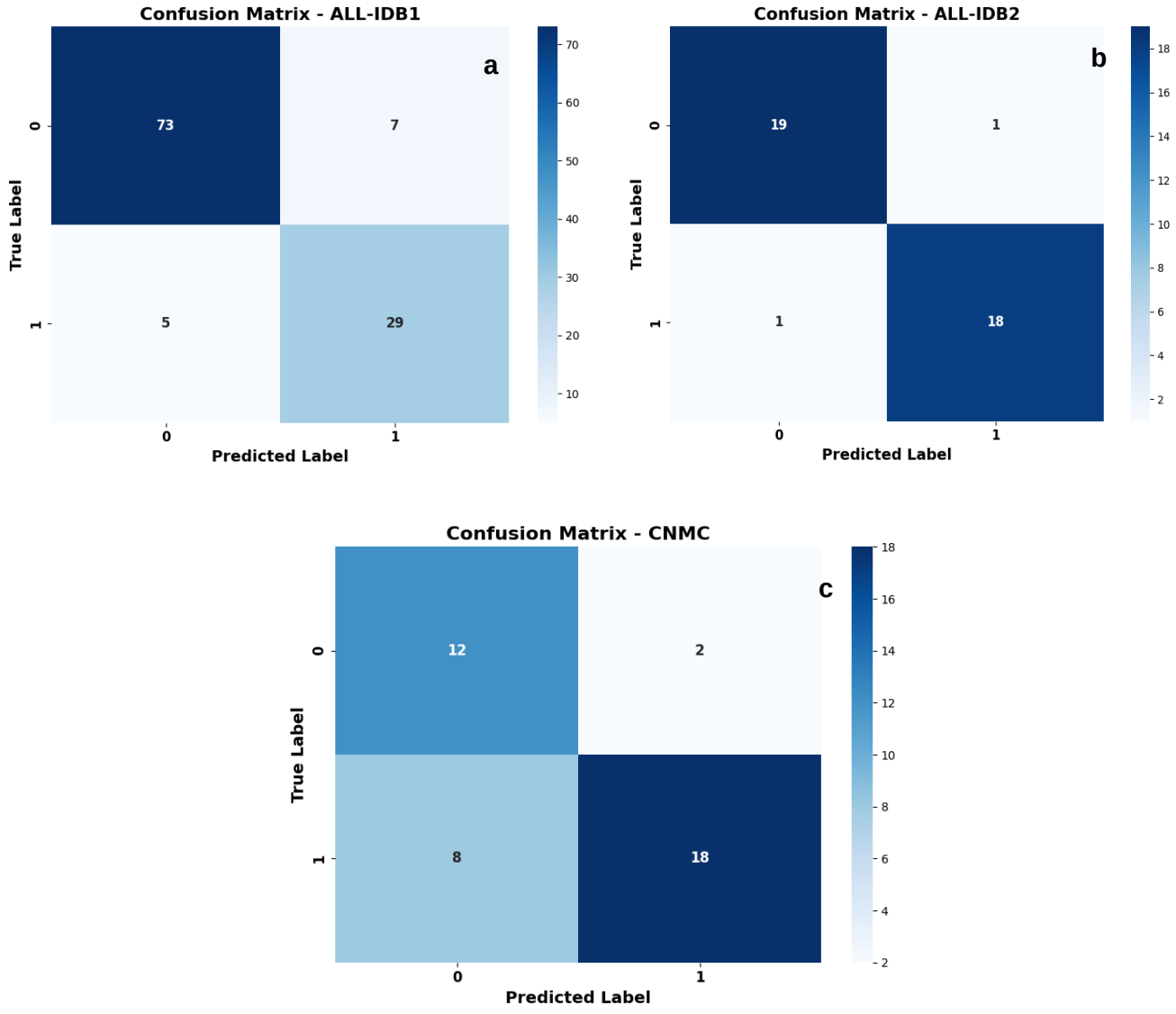

Figure 10 a). Confusion matrix: Evaluation of YOLOv8s unified model on ALL-IDB1 dataset, **b)** ALL-IDB2 dataset, **c)** C-NMC-19 dataset

YOLOv5, being anchor based with an older CSPDatrknet backbone, shows competitive performance on single-cell datasets but lacks the adaptability and architectural enhancements of YOLOv8s and YOLO11s, leading reduced generalization in diverse or unified dataset [8, 9].

The goal of generalized model is to achieve robust performance not only on training dataset, but also on unseen or heterogeneous datasets. Generalization was specifically evaluated by training YOLO models (YOLOv5s, YOLOv8s, and YOLO11s) on unified dataset and testing them on individual datasets. We evaluated the generalized YOLO models against dataset specific models by testing them on the same set of unseen images from all three datasets and compared their performances as presented in Table 7 and 8.

On ALL-IDB1 dataset, the generalized YOLOv8 (86.9% mAP@50) outperforms compared to all other models. On ALL-IDB2, generalized YOLOv8 (98.1% mAP@50) is very close to dataset-specific YOLOv8 (98.8% mAP@50).

On C-NMC_19 generalized YOLOv8 (91.1% mAP@50) perform slightly lower than dataset-specific YOLOv8 (95% mAP@50), but still better than generalized YOLOv11 and YOLOv5 models.

While generalized YOLOv8 model achieves strong performance across unseen dataset, slight performance drops compared to dataset-specific YOLOv8 model occur for C_NMC-19, primarily due to domain shift. Nevertheless, generalized YOLOv8 outperforms all other generalized models and achieves competitive results on all datasets, confirming its adaptability for cross-domain generalization while maintaining high accuracy.

In addition to accuracy and precision, inference speed is crucial factor when evaluating models for real-time deployment. As shown in table 9, YOLOv8 achieved an average inference time of 5.8ms per image, resulting in a throughput of approximately 172.41 FPS.

This value significantly exceeds the standard real-time benchmark of 30 FPS [8], demonstrating YOLOv8 is well suited for real-time application.

Figure 10a, b, c illustrates the confusion metrics of YOLOv8s Unified model evaluated across three distinct datasets providing a clear view of classification performance. Result of validation batch of YOLOv8s unified model on ALL-IDB1 actual labels and predicted labels is shown in figure 11a and b.

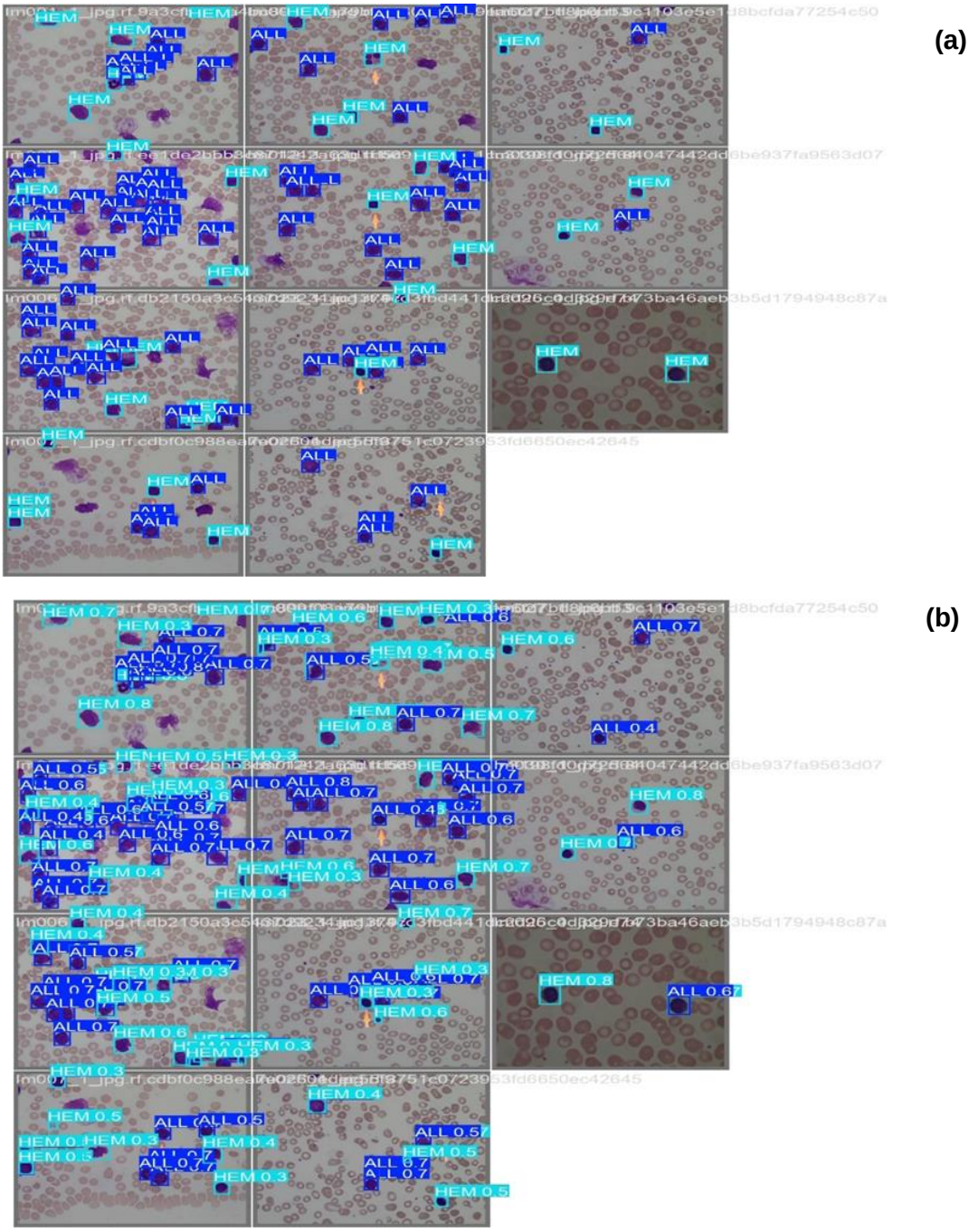

Figure 11 a. Validation batch of YOLOv8s unified model on ALL-IDB1: Actual labels, **b)** Predicted labels

Table 10. Performance comparison of the proposed generalized model- YOLOv8 with existing approaches for ALL detection

Dataset	Author	Accuracy	Generalization Capability.
C-NMC-19	Lamberti *et al.* (2022) [14]	Accuracy-90.10%	Dataset-specific models with limited generalization, trained and evaluated on single-cell image dataset.
	Pradeep Das *et al.* (2022) [15]	Accuracy-91.56%	
	Mantri *et al*. (2024) [27]	Accuracy-96.87%	
	Proposed model	**mAP-91.1% Recall-94.3%**	**Generalized model.**
ALL_IDB2	Lamberti *et al*. (2022) [14]	Accuracy-100%	Models optimized for individual datasets. Lack of Generalization. High accuracy indicates risk of overfitting.
	Pradeep Das *et al.* (2022) [15]	Accuracy-98.21	
	Proposed model	**mAP- 98.1%. Recall-97.2%**	**Generalized model.**
ALL_IDB1	Pradeep Das *et al.* (2022) [15]	Accuracy-98.21%	Dataset-specific models, trained and evaluated on multi-cell image dataset. Lack of Generalization.
	Chen *et. al.* (2022) [20]	mAP YOLOv5-90.15% YOLOv6-64.4% YOLOv7-58.3%	
	Al-Bashir *et al.* (2024) [16]	Accuracy-94%	
	Proposed model	**mAP- 86.9%. Recall-93.1%**	**Generalized model.**

Table 10 provides the comparison between proposed generalized model with existing approaches in terms of accuracy and generalization capability.

5. Conclusion

This study used a uniform dataset compiled from three diverse sources—ALL-IDB1, ALL-IDB2, and C-NMC-19—to propose a generic framework based on YOLO for the real-time detection of leukemia cells. The objective was to evaluate whether a unified training strategy can enhance model generalization across diverse imaging conditions like single cell and multi-cell datasets. We trained the each YOLO model-YOLOv5s, YOLOv8s and YOLO11s, on both individual dataset and the unified dataset. Experimental results demonstrated that the YOLOv8s consistently outperformed other variants in all cases. Comparing generalization capability, YOLOv8s achieves competitive performance against dataset-specific models, with mAP@50 reaching up to 98.1% on ALL_IDB2 and 94.3 % recall on C-NMC-19. The superior mAP@50 of generalized YOLOv8s model on ALL-IDB1 (86.9%), confirms that generalization through unified training can improve accuracy reduce overfitting and strengthen the model. For C-NMC-19 dataset, although slight dip is observed compared to data-specific model primarily due to domain shift, the performance remains strong. Furthermore, the all three model achieved inference speeds exceeding 30 FPS, meeting real-time requirements of practical applications. These findings suggests that YOLOv8 strikes the best balance between detection accuracy and generalization across diverse datasets.

Since we have limited dataset of multi- cell images i.e. ALL-IDB1, future work will focus on expanding training with more diverse datasets, increasing epochs for better generalization, fine tuning YOLO11 for Hem class detection, and validating the framework in clinical settings to guarantee dependability and relevance in practical diagnosis.

References

[1] M. Zolfaghari, H. Sajedi, A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells. Multimedia Tools and Applications, 81(5), (2022) 6723-6753. https://doi.org/10.1007/s11042-022-12108-7

[2] K. Sathishkumar, M. Chaturvedi, P. Das, S. Stephen, P. Mathur, Cancer incidence estimates for 2022 & projection for 2025: result from National Cancer Registry Programme, India. Indian Journal of Medical Research, 156(4&5), (2022) 598-607. https://doi.org/10.4103/ijmr.ijmr_1821_22

[3] R. Mantri, R.A.H. Khan, D.T. Mane, An Efficient System for Detection and Classification of Acute Lymphoblastic Leukemia Using Semi-Supervised Segmentation Technique. International Research Journal of Multidisciplinary Technovation, 7(2), (2025)121-134.

https://doi.org/10.54392/irjmt25210

[4] J. Chae, J. Kim, An Investigation of Transfer Learning Approaches to Overcome Limited Labelled Data in Medical Image Analysis. Applied Sciences, 13(15), (2023) 8671. https://doi.org/10.3390/app13158671

[5] S.S.A. Zaidi, M.S. Ansari, A. Aslam, N. Kanwal, M. Asghar, B. Lee, A survey of modern deep learning-based object detection models. Digital Signal Process, 126, (2022) 103514. https://doi.org/10.1016/j.dsp.2022.103514

[6] L. Du, R. Zhang, X. Wang, Overview of two-stage object detection algorithms, In Journal of physics: Conference series, 1544(1), (2020) 012033. https://doi.org/10.1088/1742-6596/1544/1/012033

[7] T. Mustaqim, C. Fatichah, N. Suciati, Deep learning for the detection of acute lymphoblastic leukemia subtypes on microscopic images: A systematic literature review. IEEE Access, 11, (2023) 16108-16127. https://doi.org/10.1109/ACCESS.2023.3245128

[8] J. Terven, D.M. Córdova-Esparza, J.A. Romero-González, A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. Machine Learning and Knowledge Extraction, 5(4), (2023) 1680-1716. https://doi.org/10.3390/make5040083

[9] Glenn Jocher, Paula Derrenger, and Muhammad Rizwan Munawar. Home - Ultralytics YOLO Docs, Accessed: [Online]. Available: https://docs.ultralytics.com/

[10] P. Hidayatullah, N. Syakrani, M.R. Sholahuddin, T. Gelar, R. Tubagus, (2025) YOLOv8 to YOLO11: A comprehensive architecture in-depth comparative review. arXiv preprint arXiv:2501.13400.

[11] R.D. Labati, V. Piuri, F. Scotti, (2011) All-IDB: The acute lymphoblastic leukemia image database for image processing. 18th IEEE international conference on image processing, IEEE, Belgium. https://doi.org/10.1109/ICIP.2011.6115881

[12] A. Gupta, R. Gupta, ALL Challenge Dataset of ISBI 2019 [Data Set]. The Cancer Imaging Archive, (2019). https://www.cancerimagingarchive.net/collection/c-nmc-2019/

[13] M. Zolfaghari, H. Sajedi, (2023) A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells. arXiv.

[14] W.F. Lamberti, (2022) Classification of white blood cell leukemia with low number of interpretable and explainable features. arXiv preprint arXiv, 2201.11864.

[15] P.K. Das, B. Sahoo, S. Meher, An efficient detection and classification of acute leukemia using transfer learning and orthogonal softmax layer-based model. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20(3), (2022) 1817-1828. https://doi.org/10.1109/TCBB.2022.3218590

[16] A.K. Al-Bashir, R.E. Khnouf, L.R. Bany Issa, Leukemia classification using different CNN-based algorithms-comparative study. Neural Computing and Applications, 36(16), (2024) 9313-9328. https://doi.org/10.1007/s00521-024-09554-9

[17] S. Raina, A. Khandelwal, S. Gupta, A. Leekha, (2020) Blood cells detection using faster-RCNN. GUCON, IEEE International Conference on Computing, Power and Communication Technologies (GUCON), IEEE, India. https://doi.org/10.1109/GUCON48875.2020.9231134

[18] A.R. Revanda, C. Fatichah, N. Suciati, Classification of Acute Lymphoblastic Leukemia on White Blood Cell Microscopy Images Based on Instance Segmentation Using Mask R-CNN. International Journal of Intelligent Engineering & Systems, 15(5), (2022) 625-637. https://doi.org/10.22266/ijies2022.1031.54

[19] R. Khandekar, P. Shastry, S. Jaishankar, O. Faust, N. Sampathila, Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis. Biomedical Signal Processing and Control, 68 (2021) 102690. https://doi.org/10.1016/j.bspc.2021.102690

[20] E. Chen, R. Liao, M.Y. Shalaginov, T.H. Zeng, Real-time detection of acute lymphoblastic leukemia cells using deep learning, 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, USA. https://doi.org/10.1109/BIBM55620.2022.9995131

[21] C.K. Chou, R. Karmakar, Y.M. Tsao, L.W. Jie, A. Mukundan, C.W. Huang, T.H. Chen, C.Y. Ko, H.C. Wang, Evaluation of spectrum-aided visual enhancer (SAVE) in esophageal cancer detection using YOLO frameworks. Diagnostics, 14(11), (2024) 1129. https://doi.org/10.3390/diagnostics14111129

[22] S. Kundu, A. Dutta, K.K. Jha, (2024) Analysis and Identification of Leukemia Using YOLOv8, In 2024 4th International Conference on Computer, Communication, Control & Information Technology (C3IT), IEEE, India. https://doi.org/10.1109/C3IT60531.2024.10829424

[23] J. Redmon, (2016) You only look once: Unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, USA. https://doi.org/10.1109/CVPR.2016.91

[24] G. Jocher, (2020) Ultralytics YOLOv5. GitHub.

https://github.com/ultralytics/yolov5

[25] G. Jocher, J. Qiu, (2024) Ultralytics YOLO11, version 11.0.0, https://github.com/ultralytics/ultralytics

[26] G. Jocher, A. Chaurasia, J. Qiu, (2023) Ultralytics YOLOv8. Ultralytics, https://github.com/ultralytics/ultralytics

[27] R. Mantri, Khan, R.A.H., Jadhav, S.;Leukemia Diagnosis using Transfer Learning: An Efficient Approach, Frontiers in Health Informatics, 13(2) (2024) 481-496.

Authors Contribution Statement

Ratnamala Mantri (Paswan): Conceptualization, Dataset preparation. Methodology, Model analysis, Writing-Review & Editing. Rais Abdul Hamid Khan: Conceptualization, Dataset verification, Model analysis, Model verification, Supervision. Both the authors have read and agreed to the published version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes

About the License