YOLOV5 : Classify Kaggle WBC data set using transfer learning approach using LISC data set.

Vandita Sharma[1] and Dr. Tilak Raj Rohilla [2]

[1] *Research Scholar, Baba MastNath University, Asthal Bohar, Rohtak Haryana, India*
[2] *Assistant Professor, Baba MastNath University, Asthal Bohar, Rohtak Haryana, India*
Sharma21vandita@gmail.com
Tilak.rohilla@gmail.com

KEYWORDS

Deep Learning, YOLOv5, Transfer Learning, WBC

ABSTRACT

White blood cells (WBCs), are essential constituents of the immune system, by providing organism's defence against infections, inflammation, and various diseases. Types of WBCs, possessing specific functions and attributes that are vital for the maintenance of health. The categories of WBCs are neutrophils, lymphocytes, monocytes, eosinophils, and basophils, each performing functions in the boost immune response. YOLOv5 stands out as a top-tier object detection model recognized for its speed and accuracy in detecting objects from images. The YOLOv5 model serves as an effective tool for detecting and classifying different white blood cell types in blood smear images during WBC classification and identification tasks. In Transfer learning A model developed for a specific task is reused with change in hyperparameters, as the starting point for a model on a second task. This approach is particularly valuable in scenarios where limited labelled data is available for the target task, allowing practitioners to leverage knowledge gained from related tasks with ample data. Transfer learning is commonly used in deep learning, especially for tasks such as image classification, natural language processing, and speech recognition. Kaggle and LISC original data set has limited number of images and YOLOv5 is state of art model for object detection and classification, and the employing transfer learning is the core idea of this research. Learning features from Original LISC data set and applying best weights produced from previous run we applied that weight on Augmented images of Kaggle data set and produced accuracy of 99.5mAP@50, recall 99 and F1-Score 99.

Introduction:

Blood specimens is indispensable for diagnosing a multitude of health afflictions, encompassing infections, hypersensitivities, and haematological disorders. The procedure entails recognizing and classifying various types of white blood cells present in a blood smear, which can yield significant insights into a patient's immune response and general health. Preparation of a blood smear on a glass slide, which is stained to augment the visibility of the cells. Automated systems, such as digital imaging and machine learning algorithms, are progressively employed to detect WBCs in these stained visuals. Methodologies such as image processing and computational vision are utilized to locate the cells, frequently employing techniques such as edge detection and contour analysis to precisely delineate the boundaries of the cells. Each type of WBC assumes a distinct role within the immune system, and their relative proportions can signify specific health conditions. Classification algorithms, such as convolutional neural networks (CNNs), characteristics of the detected cells and assign

designations. These models are trained on extensive datasets of annotated blood smear images to enhance their precision and dependability.
Deep learning, a subdivision of Machine Learning, entails the utilization of neural networks characterized by numerous layers (thus the term "deep") to emulate intricate patterns within substantial datasets. Deep learning and human brain work on same principle, employing interconnected nodes (neurons) to process information. Each layer within the network extracts progressively abstract features from the input data. Employing this in the domain of image recognition, works as the initial layer may discern edges, the subsequent layer identifies shapes, and deeper layers may recognize objects. This hierarchical feature extraction enables deep learning models to attain high level of feature extraction and accuracy in tasks such as image classification [1]. Detection and classification and computer vision often use machine, deep learning, in applications such as image processing, object recognition, and natural language processing. Detection methodology ascertain the existence of particular entities or configurations within an image or a dataset. It entails the pinpointing and acknowledgment of objects of significance and denote the locations of the identified objects e.g., Bounding Box etc.. YOLO [2] (You Only Look Once) and Faster R-CNN [3], scrutinize an image and producing the coordinates of recognized objects alongside their respective classifications. Detection is pivotal in surveillance, autonomous vehicles, and medical imaging, wherein the precise identification and localization of objects is imperative. Classification, conversely, encompasses the allocation of a label or category to an entire image or a segment of data predicated on its content. Both detection and classification, discerns the existence of objects but also categorizes them can be executed concurrently in tasks such as object detection. The amalgamation of these tasks enhances the capacity of systems to comprehend and interpret intricate data, culminating in more resilient applications in real-world situations. [4]

White blood cells (WBCs), are essential constituents of the immune system, by providing organism's defence against infections, inflammation, and various diseases. Types of WBCs, possessing specific functions and attributes that are vital for the maintenance of health. The categories of WBCs are neutrophils, lymphocytes, monocytes, eosinophils, and basophils, each performing functions in the boost immune response. Neutrophils responds to areas of infection or injury. Lymphocytes, which comprise T & B cells, are indispensable for adaptive immunity; T cells assist in immune responses and assault infected cells, while B cells synthesize antibodies to neutralize pathogens. Monocytes differentiate resident tissue and professional antigen-presenting cells, significant for phagocytosis and antigen presentation. In allergic reactions and responses to parasitic infections, Eosinophils and Basophils partake. Eosinophils primarily targeting larger parasites and basophils releasing histamine during inflammatory responses. An elevated WBC count (leukocytosis) may signify infection or stress, while a diminished count (leukopenia) can imply bone marrow disorders or autoimmune diseases. Differential WBC Count offers a aggregation of the neutrophils, lymphocytes, monocytes, eosinophils, and basophils. Each type of WBC possesses distinct functions in the immune response, and their proportions can facilitate the diagnosis of specific infections, allergies, and hematological malignancies. Platelet quantifies the number of platelets and essential for hemostasis and the prevention of excessive hemorrhage. Aberrant platelet counts can indicate bleeding disorders (thrombocytopenia) or thrombotic conditions (thrombocytosis). [5].

YOLOv5[26] stands out as a top-tier object detection model recognized for its speed and accuracy in detecting objects from images. The YOLOv5 model serves as an effective tool for detecting and classifying different white blood cell types in blood smear images during WBC classification and identification tasks. This model operates by dividing the image into a grid structure and predicts bounding boxes along with class probabilities for each grid cell to enable real-time detection and categorization of multiple objects within images.

Transfer learning approach is particularly valuable in scenarios where limited labeled data is available for the target task, allowing practitioners to leverage knowledge gained from related tasks with ample data. Transfer learning involves training model on large dataset and fine tune it on targeted domain specific dataset. this way we can reduce the training time and at the same time can increase performance of the model. In computer vision tasks, models like VGGNet [6], ResNet [7], and Inception [8] are often pre-trained on large datasets such as ImageNet and then adapted for specific applications such as medical image analysis or object detection.

YOLOv5 has benefits e.g., Real time and Multiclass classification, integration with other techniques, domain specific, widely acceptable for clinical applications. YOLOv5 is recognized for its instantaneous object detection capabilities, which renders it particularly appropriate for clinical applications where expeditious analysis is paramount. YOLOv5 has been modified to categorize multiple varieties of WBCs concurrently. Researchers have formulated datasets that encompass an array of WBC types, such as neutrophils, lymphocytes, monocytes, eosinophils, and basophils, permitting the model to assimilate distinguishing features for each category. Certain investigations have emerged in which YOLOv5 is used with Convolutional Neural Networks (CNNs) or Generative Adversarial Networks (GANs), Transformers etc. to bolster the robustness of WBC classification. Metrics such as precision, recall, F1-score, and mean Average Precision (mAP) are typically used for comprehensive evaluation of the models. Automated systems employing YOLOv5 for WBC detection and classification can assist pathologists by supplying preliminary analyses and diminishing the time necessitated for manual scrutiny of blood smears. [9][10][11].
Literature Survey: The table compares various YOLO-based models for blood cell detection and classification across different datasets, highlighting their architectures, computational efficiency, and performance metrics. [12] introduces a modified YOLOv3 using EfficientNet-B3 and depth-wise separable modules on the BCCD dataset, achieving a micro-average precision (µAP) of 89.86% with a 5x reduction in parameters and 0.57 seconds detection time. Similarly, [25] proposes TE-YOLOF with Mish activation and EfficientNet-B3, attaining a higher µAP of 91.9% but with increased parameters. [26] employs a two-phase YOLOv3 approach on BCCD, achieving 99% detection and 90% classification accuracy. [27] applies YOLOv5 with CSP modules and transfer learning on the Raabin dataset, achieving 94.66% accuracy, with detection times varying by model size (0.5–2.7 hours per epoch). For feature extraction, [28] combines YOLOv3 with AlexNet on the LISC dataset, achieving 98% accuracy but with slower detection (74.4 seconds). [29] reports 88% accuracy for YOLOv5 on a public dataset, while [31] achieves higher accuracy (89.25%) and recall (96.9%) on a custom dataset. Tiny YOLO [32] and YOLOv4 ([33] on BCCD show comparable accuracy (~86.89%), but Tiny YOLO excels in recall (99%) and F1-score (98%). [36] improves YOLOv4's performance on BCCD, achieving 97.8% accuracy and 95.7% F1-score. Notably, models like YOLOv5 [27, 31] and modified YOLOv3 [24] demonstrate efficiency and accuracy trade-offs, while deeper architectures (e.g., YOLOv4 in [36]) excel in precision but require more computational resources.

Ref.	Model Based	Data Set	Key characteristics	Detection Time	Performance Matrices
[12]	Modified YOLOv3	BCCD	BB: EfficientNet-B3, Dilated. AF: Swish LF: DIoU. Head: Depth wise separable Module. Number of Parameters Reduced by 5x	0.57 Sec.	Platelets: 90.25 RBC: 80.41 WBC: 98.92 µAP=89.86
[13]	TE-YOLOF	BCCD	BB: Efficient Net-B3 Head: Depth wise separable Module. AF: Mish. Numbers of Parameters are very High.	Epochs-12	Platelets: 89.8 RBC: 87.3 WBC:98.7 µAP=91.9
[14]	YOLOv3	BCCD	Divided in to Two Phases: Phase1: Generation of WBC Bounding Box. Phase2: Classification and detection of WBC.	NA	Detection: 99%. Classify: 90%.
[15]	YOLOv5	Raabin Health Database	BB: 4 Different feature Maps. (132,76,38,19) Neck Network: CSP, module is formed by replacing with CBL module. Transfer learning.	Ephochs-200 YoloV5s-0.5h YoloV5x-2.7 h.	94.66%
[16]	YOLOV3 with CNN Feature extractor models.	LISC	Feature Extractor: AlexNet. AF: ReLU.	74.4sec	98%
[17]	Yolov5	Publicly available dataset	Several in-depth research strategies employed for WBC classification.	NA	88%
[19]	Tiny YOLO (2019)	BCCD (364)	Tiny Yolo.	NA	Accuracy (%) 86.89 Recall (%) 99 F1 score (%) 98

[20]	YOLOv4	BCCD (364)	YOLOv4	NA	Accuracy (%) 95.75 Recall (%) NA F1 score (%) NA
[18]	YOLOv5	855 (own dataset)	YOLOv5	NA	Accuracy (%) 89.25 Recall (%) 96.9 F1 score (%) 94
[21]	Yolov4	BCCD	YOLOv4	NA	Accuracy (%) 97.8 Recall (%) 95.7 F1 score (%) 95.7
[25]	YOLOv5	619	YOLOv5	NA	Accuracy 93.0%

Table1: Literature Survey of Model based on YOLO Based models and their Vitals

III Methodology.

In this paper our objective is to detect and classify WBCs using popular data set available on RoboFlow website, BCCD [23] and LISC [22]. To achieve this we are using YOLOV5 to classify and detect WBC. Our approach is simple yet effective, we trained YOLOV5 Variants Nano, Small, Medium, Large on original LISC data and then opted best model based on K-Fold Cross Validation. Our parameters for selection were high mAP@50 and Precision. Then we augmented Kaggle data set images and used best model obtained from first run tweaked some Hyper parameters like freezing layers and changing anchor boxes size based on the Input images size.

A. Data Augmentation.

We crated pipeline that applies a robust set of transformations to enhance model generalization by simulating real-world variations in the data. Key augmentations include Random Resized Crop, Flipping & Rotation, Shift Scale Rotate, Color & Lighting Adjustments, Noise & Distortions, Gauss Noise, Elastic Transform and Perspective. Aim of above transformations to improve robustness, avoid excessive distortion. These Augmentations targets tasks like object detection/ classification where invariance to scale, orientation, and lighting is critical.

B. Evaluation parameters of YOLOv5

Accuracy. Rates both true positives and true negatives out of the total illustrations. This measure can be misleading when the user is working on a balanced dataset. **Precision.** Quantifies how many of the predicted positive instances are true positives. This measure helps minimize false positives, which is crucial. **Recall.** Rates the proportion of actual positive instances that were correctly predicted. This measure helps when minimizing false negatives is crucial. **F1-score.** Balances precision and recall, providing a harmonic mean of the two. This measure is useful when both false positives and false negatives need consideration.

$$\text{Accuracy} = \frac{\text{Correct Prediction}}{\text{Total Prediction}} \quad (1)$$

$$\text{Precision} = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

$$\text{F1} - \text{Score} = \frac{Precision\ .Recall}{Precision + Recall} \quad (4)$$

IV. Data Set.

	Training	Validation
Initial Data (LISC)	681	171
Augmentation Data (BCCD)	9570	957

Table 2. Number of total images in initial data and augmentation data

The table presents the distribution of images across training, validation sets before and after data augmentation. Initially, the dataset contained a small number of images with 681 for training, 171 for validation, and only 17 for testing, which is insufficient for training a deep learning model effectively. To improve model generalization and performance, data augmentation was applied, significantly increasing the dataset size. After augmentation, the training set expanded to 9,570 images, while the validation and test sets increased to 957 respectively. This augmentation process ensures the model learns from a diverse set of variations, reducing overfitting and improving robustness. The augmented dataset provides a much stronger foundation for training an object detection model like YOLOv5, leading to better performance on unseen data. Goole Colab Infrastructure's Tesla T4 GPU used to run above use case. T4 instance used for this run have system RAM12.7 GB, GPU RAM 15.00 GB, Disk 112.6 GB. And we have connected this instance with Google Drive where we have kept our test data and Results.

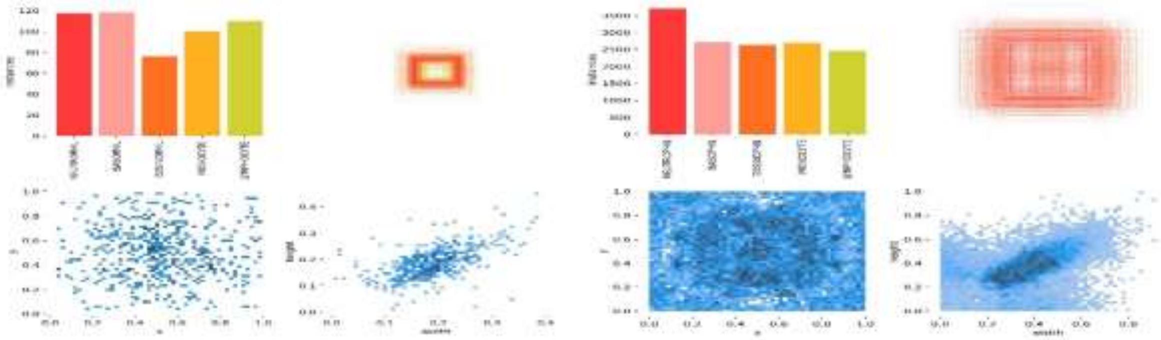

Figure3: LISC(Left) and Kaggle(Right) data Distribution used for Initial Run.

LISC dataset visualization provides crucial insights into class distribution, object locations, and bounding box sizes. The bar chart shows an imbalance, with Neutrophils and Basophils appearing more frequently, while Eosinophils have fewer instances, which could contribute to detection challenges. The bounding box heatmap and object location heatmap reveal that most objects are centred within the images, with fewer detections at the edges. The width-height scatter plot confirms that most bounding boxes are small to medium-sized, indicating relatively uniform object sizes. Kaggle dataset analysis highlights important patterns in class distribution, object locations, and bounding box sizes. The bar chart shows a relatively balanced class distribution, though Neutrophils are still dominant, which may lead to class imbalance issues. The bounding box heatmap reveals that most objects are centred, suggesting a potential bias toward middle placements. The object location heatmap confirms that objects are well spread but slightly concentrated at the centre. Finally, the bounding box width-height scatter plot indicates a consistent object size distribution, with most objects following a proportional aspect ratio. Kaggle data set insights suggest possible improvements such as data augmentation to diversify object placements, rebalancing classes to prevent bias, and refining anchor boxes to better match varying object sizes. LISC data set insights suggest potential improvements such as data augmentation for minority classes, cropping or resizing strategies to balance object placement, and fine-tuning anchor boxes to better capture object scales.

The YOLOv5 training curves indicate a well-converging model with strong performance. The training losses—including box loss (bounding box regression), objectness loss (confidence in detecting objects), and classification loss (object classification accuracy)—are all steadily decreasing, demonstrating that the model is effectively learning to localize and classify objects over time. Similarly, the validation losses follow a decreasing trend, confirming that the model is generalizing well to unseen data without signs of overfitting.

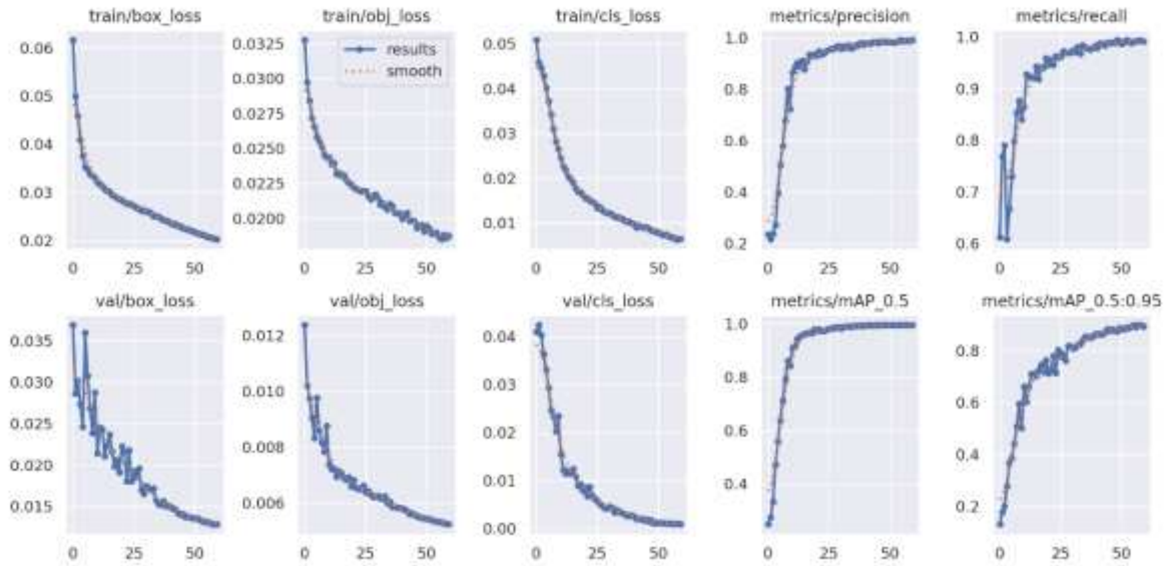

Figure 4. The results of the classification of blood cells

The precision and recall metrics show a rapid improvement in the early epochs, with precision stabilizing near 1.0, indicating very few false positives, and recall steadily increasing, reflecting a high detection rate. The mean Average Precision (mAP) at IoU=0.5 (mAP_0.5) reaches around 0.99, showcasing excellent detection accuracy. Meanwhile, the mAP across multiple IoUs (mAP_0.5:0.95) improves to approximately 0.85, which is a strong indicator of robust generalization across different object sizes and shapes.

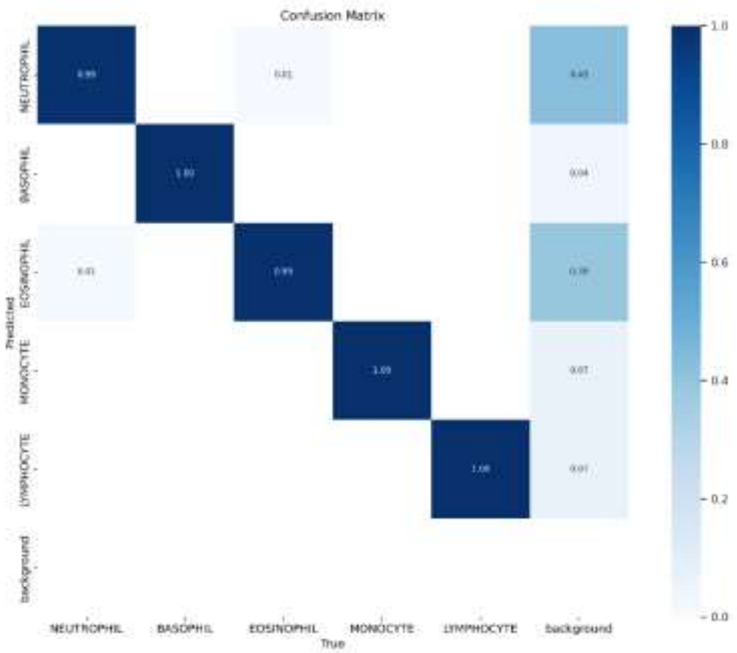

Figure 4. The results of the classification of blood cells

The updated confusion matrix shows a significant improvement in classification accuracy for blood cell detection. The model demonstrates near-perfect performance, correctly identifying Neutrophils (99%), Basophils (100%), Eosinophils (99%), Monocytes (100%), and Lymphocytes (100%) with minimal misclassification. The only minor errors occur between Neutrophils and Eosinophils, with a 1% misclassification rate between these two classes. However, background misclassification remains a concern, with 43% of background instances being incorrectly classified as Neutrophils, 39% as Eosinophils, and 7% each as Lymphocytes or Monocytes. While the classification of blood cells has significantly improved, further refinements—such as better background sampling or incorporating more negative samples—may help reduce background misclassification and enhance overall model performance.

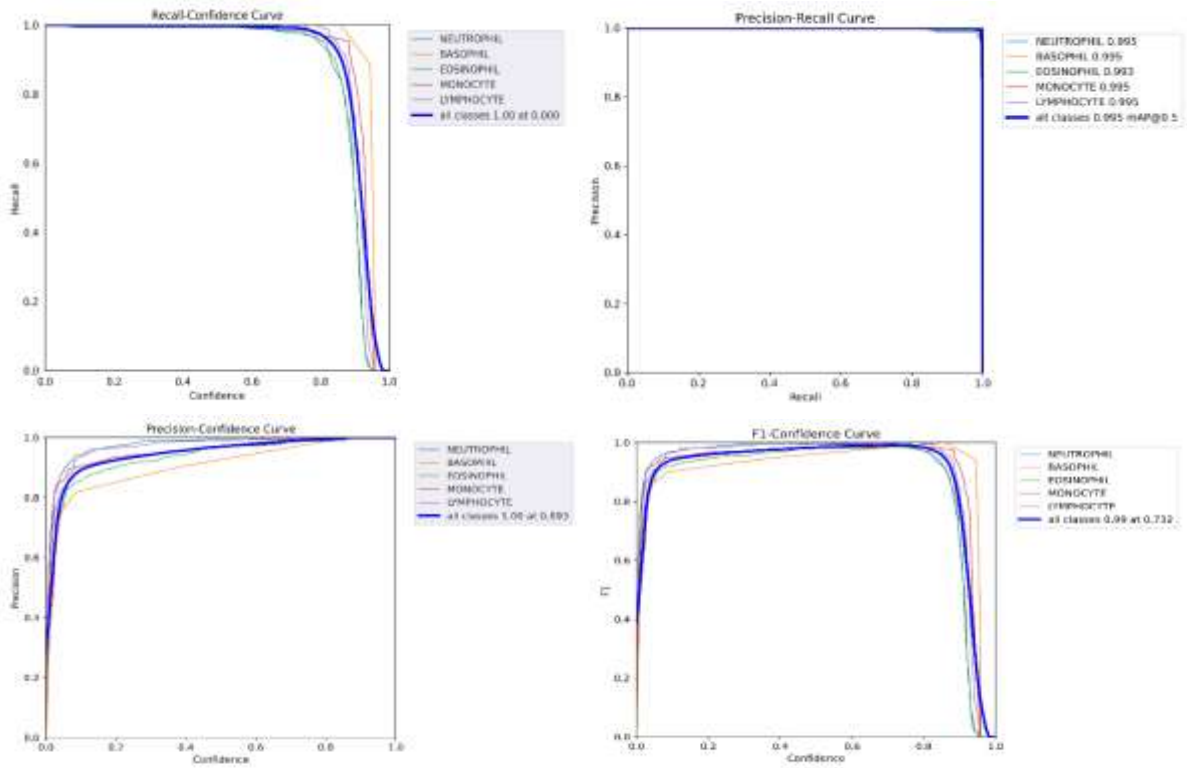

Figure5: RC(Top Left), PR (Top Right), PC(Bottom Left) and F1- curves(Bottom Right). RC: Recall-Confidence, PR: Precision-Recall, PC (Precision- Confidence).

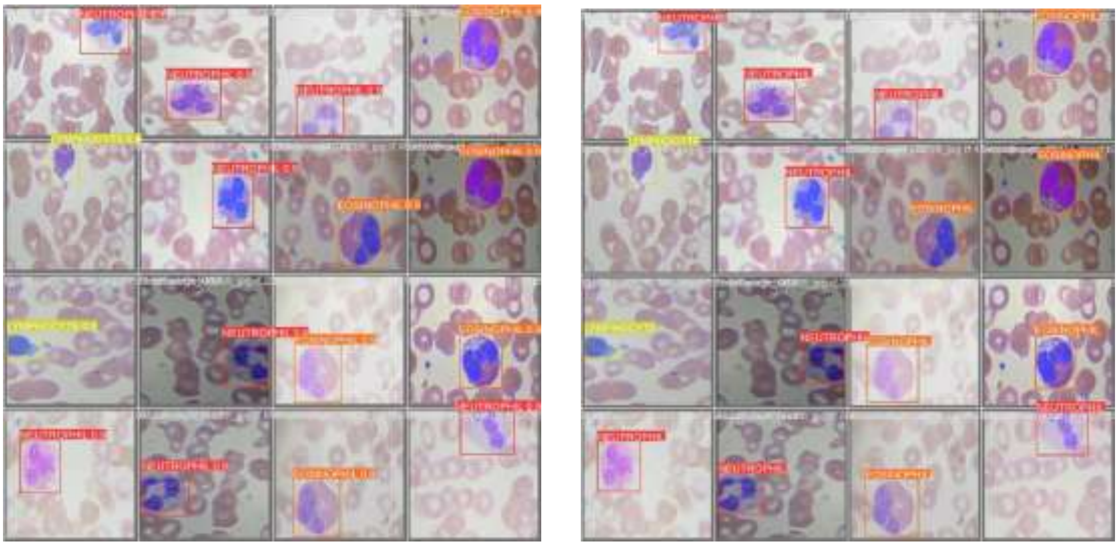

Figure6: Predicted Images mosaic (Left) and Actual labels Images mosaic (Right).

VI. Discussion

Work	Year	No. Images	Sources	Method	overall Accuracy.	Precision	Recall
[27]	2019	874	BCCD	YOLOV5 YOLOV4 YOLOV3	91.00 89.00 86.00	71.00 82.00 84.00	NA
[19]	2019	364	BCCD	Tiny YOLO	98	86.89	99
[20]	2021	364	BCCD	YOLOv4	NA	95.75	NA
[18]	2022	855	BCCD	customized YOLOv5 m	99.4	99.4	96.9
[24]	2023	10275	-	YOLOv3 with Alexnet (2020)	98	NA	NA
Proposed work	2025	9570	LISC->Kaggle	YOLOV5	99.2	99	100

Table 4 : Comparison of Accuracy, Precision and F1 Score of our method with other models.

[34] Used a customized YOLOv5 model 855 BCCD images to achieve the highest reported accuracy (99.4%), precision (99.4%), and recall (96.9%). Tiny YOLO in [32] used 364 BCCD images reached 98% accuracy but lower precision (86.89%). The proposed work uses combination of 9,570 LISC/Kaggle images outperformed most with YOLOv5, attaining 99.2% accuracy, 99% precision, and 100% recall. Notably, larger datasets [35] did not always correlate

with higher performance, and precision/recall values were missing in [27], [29], [33]. The trend highlights YOLOv5's adaptability and the impact of model customization on performance.

Conclusion:

Overall, these results confirm that the transfer learning approach and training strategy were effective, resulting in a model that performs accurately and efficiently. Further improvements could focus on using a larger model (YOLOv5m/l) for potentially better performance on stricter evaluation metrics (mAP_0.5:0.95). Other key improvements can be combining images from various established WBC data set and then training model with those images will provide more generalization to the model. The current model is well-trained, achieving a strong balance between precision, recall, and detection accuracy. Overall, these results indicate that YOLOv5S is highly effective in blood cell detection, with excellent precision and strong mAP scores.

Reference

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015, June 8). You only look once: Unified, Real-Time Object Detection. arXiv.org. https://arxiv.org/abs/1506.02640
[3] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2013, November 11). Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv.org. https://arxiv.org/abs/1311.2524.
[4] Szeliski, R. (2010). Computer vision: Algorithms and applications. Springer.
[5] McPherson, R. A., & Pincus, M. R. (2017). Henry's clinical diagnosis and management by laboratory methods (23rd ed.). Elsevier.
[6] Simonyan, K., & Zisserman, A. (2014, September 4). Very deep convolutional networks for Large-Scale image recognition. arXiv.org. https://arxiv.org/abs/1409.1556
[7] He, K., Zhang, X., Ren, S., & Sun, J. (2015, December 10). Deep residual learning for image recognition. arXiv.org. https://arxiv.org/abs/1512.03385

[8] Barratt, S., & Sharma, R. (2018, January 6). A note on the inception score. arXiv.org. https://arxiv.org/abs/1801.01973
[10] Kaur, M., & Kumar, A. (2021). A novel approach for white blood cell classification using YOLOv5. Journal of King Saud University - Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2021.05.008
[11] Sahu, S. K., & Sahu, A. (2021). A review on white blood cell detection and classification using deep learning techniques. International Journal of Medical Engineering and Informatics, 13(1), 1-15. https://doi.org/10.1504/IJMEI.2021.114885
[12] A. Shakarami, M. B. Menhaj, A. Mahdavi-Hormat, and H. Tarrah, "A fast and yet efficient YOLOv3 for blood cell detection," *Biomedical Signal Processing and Control*, Apr. 01, 2021.
[13]F. Xu, X. Li, H. Yang, Y. Wang, and W. Xiang, "TE-YOLOF: Tiny and efficient YOLOF for blood cell detection," *Biomedical Signal Processing and Control*, Mar. 01, 2022.
[14] N. Praveen, N. S. Punn, S. K. Sonbhadra, S. Agarwal, M. Syafrullah, and K. Adiyarta, "White blood cell subtype detection and classification," Oct. 20, 2021.
[15] F. Akalin, "Detection and classification of white blood cells with an improved deep learning-based approach," *TÜBİTAK Academic Journals*.
[16] N. Rohaziat, M. N. Razali, W. Nurshazwani, and N. Othman, "White Blood Cells Detection using YOLOv3 with CNN Feature Extraction Models," *International Journal of Advanced Computer Science and Applications*, Jan. 01, 2020.

[17] Sharma, V. et al. (2024). Optimization on Yolov5 to Improve Accuracy for Classification of White Blood Cells. In: Tavares, J.M.R.S., Rodrigues, J.J.P.C., Misra, D., Bhattacherjee, D. (eds) Data Science and Communication. ICTDsC 2023. Studies in Autonomic, Data-driven and Industrial Computing. Springer, Singapore.
[18] Escobar, F. I. F., Alipo-On, J. R. T., Novia, J. L. U., Tan, M. J. T., Karim, H. A., & AlDahoul, N. (2023). Automated counting of white blood cells in thin blood smear images. *Computers & Electrical Engineering*, *108*, 108710. https://doi.org/10.1016/j.compeleceng.2023.108710
[19] M. M. Alam and M. T. Islam, "Machine Learning Approach of Automatic Identification and Counting of Blood Cells," Healthcare Technology Letters 6, no. 4 (2019): 103–108, https://doi.org/10.1049/htl.2018.5098.
[20]. J. Yao, X. Huang, M. Wei, et al., "High-Efficiency Classification of White Blood Cells Based on Object Detection," Journal of Healthcare Engineering 2021 (2021): 1615192, https://doi.org/10.1155/2021/1615192.
[21] Trong, L. D., Thanh, T. P., Manh, H. P., & Minh, D. N. (2022). Yolov4 in white blood cell classification. In *Algorithms for intelligent systems* (pp. 387–400). https://doi.org/10.1007/978-981-19-6004-8_31
[22] WBCs LICS dataset. https://universe.roboflow.com/wbcs/wbc-lisc. last accessed 12 Mar 2024
[23] BCCD WBC dataset. https://public.roboflow.com/object-detection/bccd/4. last accessed 13 Mar 2024.
[24] Rohaziat, N., Razali, M., Nurshazwani, W., & Othman, N. (2020). White Blood Cells Detection using YOLOv3 with CNN Feature Extraction Models. *International Journal of Advanced Computer Science and Applications*, *11*(10). https://doi.org/10.14569/ijacsa.2020.0111058.
[25] D.T. Luong, D.D. Anh, T.X. Thang, H.T.L. Huong, T.T. Hanh, D.M. Khanh. "Distinguish normal white blood cells from leukemia cells by detection, classification, and counting blood cells using YOLOv5". In IEEE 7th National Scientific Conference on Applying New Technology in Green Buildings (AtiGB), pp. 156-160, 2022.
[26] Glenn, J., & Wang, D. (2021). YOLOv5: A state-of-the-art object detection model. Journal of Biomedical Informatics, 117, 103800. https://doi.org/10.1016/j.jbi.2021.103800