IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Acute Lymphoblastic Leukemia Detection Using Improved YOLOv5

Nishita BS

Department of Information Science and Engineering, Bangalore Institute of Technology *Bengaluru, Karnataka*

Dr.Shilpa M

Department of Information Science and Engineering ,Bangalore Institute of Technology *Bengaluru, Karnataka*

Fahad Siddiqui

Department of Information Science and Engineering, Bangalore Institute of Technology *Bengaluru, Karnataka*

Ayush Saxena

Department of Information Science and Engineering, Bangalore Institute of Technology *Bengaluru, Karnataka*

Karuna Vernekar

Department of Information Science and Engineering, Bangalore Institute of Technology *Bengaluru, Karnataka*

Abstract: This study focuses on the detection of Acute Lymphoblastic Leukemia (ALL) using an improved YOLOv5 model. Leukemia, a common form of blood cancer, disrupts normal blood cell production, emphasizing the need for accurate detection methods for early diagnosis and treatment. The study explores the use of deep learning techniques, specifically YOLOv5, for classifying leukemia and healthy cells from microscopic blood cell images. Data augmentation techniques were employed to enhance the dataset, and the model's performance was compared with traditional classifiers. Results show that the YOLOv5 model enhanced with SENet achieved the highest accuracy of 90% in classifying leukemia cells. The study highlights the potential of deep learning in improving leukemia diagnosis and treatment outcomes.

Keywords: Acute Lymphoblastic Leukemia, YOLOv5, Deep Learning, Data Augmentation, Microscopic Blood Cell Images, SENet, Classification, Diagnosis, Treatment, Accuracy

I. Introduction

Leukemia, a prevalent form of blood and bone marrow cancer, disrupts the normal production of blood cells, leading to a proliferation of abnormal white blood cells. Its significance stems from its widespread occurrence across age groups, impacting the immune system and overall health. Recognizing the symptoms, such as fatigue and frequent infections, is crucial for early diagnosis.

Accurate detection methods are pivotal in managing leukemia effectively. Traditional diagnostic tools like blood tests, bone marrow biopsy, and advanced genetic testing help identify specific subtypes and tailor treatment plans accordingly. Early intervention with targeted therapies, including chemotherapy and immunotherapy, significantly improves patient outcomes. Ongoing research focuses on refining detection methods to enhance precision and facilitate timely interventions, underscoring the importance of staying at the forefront of leukemia diagnostics for better patient care.

Blood cells can become contaminated with cancerous cells, which can infiltrate multiple organs and cause harm to the body [1]. If the rapid growth of abnormal cells isn't detected and treated in time, bone marrow depletion can lead to severe complications. The risk gradually decreases until the late 20s, when it begins to rise again. According to the American Cancer Society, ACS estimates 6660 cases of ALL in the US in

2022 children and adults. The ALL risk is high in children younger than five years old [2] There are different types of leukemia that haematologists in cell transplant laboratories can differentiate/diagnose based on microscopic images. If the slide is correctly stained, some types of leukemia can be more easily identified and distinguished than others, but more equipment is needed to determine underlying leukemia. Figure 1 shows the stained slides of the most common different types of leukemia.

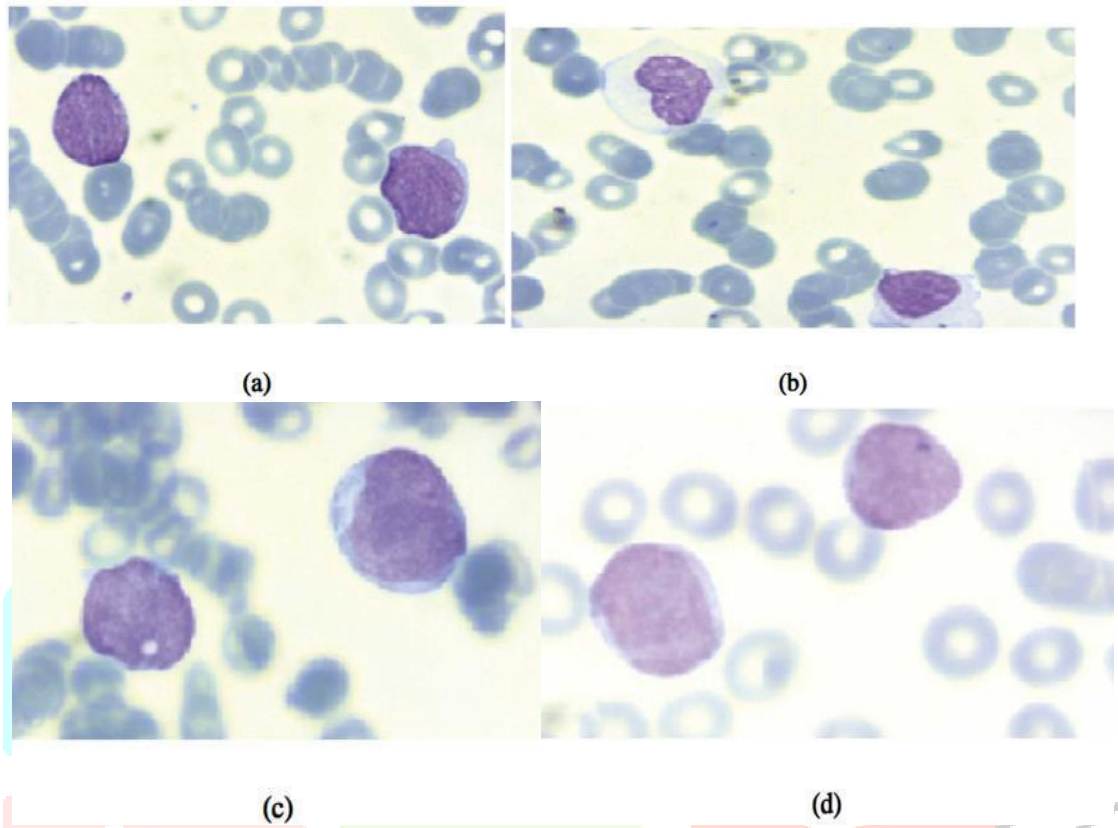


Figure 1. (a) AML (M1), (b) AML (M2), (c) B-ALL (pre-B), and (d) B-ALL (pro-B).

In this paper we shall be focusing on classifying the cells into leukemia and healthy cells.

II. Related Work

The analysis of infected blood cell images is commonly segmented into three key stages: initial image preprocessing, feature extraction, and subsequent feature selection for classification. Numerous studies have delved into various cancer types, such as leukemia, lymphoma, and myeloma. In the context of cervical cell analysis, Zhang et al. introduced a convolutional neural network model designed for the direct classification of cervical cells, distinguishing between infected and uninfected cells without the need for segmentation [3].

Table1 and Table2 and illustrate the segmentation performance using BCCD and ALL-IDB2 datasets. These tables indicate the CNN-based segmentation model suggested by Banik et al.
[4] yields superior performance in both the datasets with an accuracy of 99.42% and 98.61%, respectively. It yields good performance due to the fusion of features of the first and last convolution layers. Zhao *et al.* proposed machine learning algorithms like CNN, SVM, Random Forests, etc. for classifying various types of white blood cells (WBCs) present in the body [5]. Goswami et al. [7] have suggested a transfer learning (InceptionV3)-based ALL detection scheme in which they have emphasized optimization of heterogeneity loss that helps the network for learning of subject-independent features. This proposed work is validated using the C-NMC dataset [6], which is the largest available ALL dataset. It achieves 95.26% of the weighted F1 score. Gupta and Gupta [8] have discussed some important challenging factors of the C-NMC

dataset [6] that yields ALL detection tougher. The morphological similarity between ALL and healthy images,

Table 1. Segmentation performance using BCCD dataset.

Method	Precision (%)	Specificity (%)	Sensitivity (%)	Accuracy (%)
[152]	80.51	99.30	**94.51**	99.15
[153]	87.77	99.59	89.77	99.22
[104]	15.81	86.64	51.46	85.79
[154]	3.86	14.17	77.76	75.88
[3]	**94.38**	**99.78**	91.21	**99.42**

Table 2. Segmentation performance using ALLIDB2 dataset.

Method	Precision (%)	Specificity (%)	Sensitivity (%)	Accuracy (%)
[152]	91.24	98.62	98.09	98.59
[153]	96.00	**99.48**	88.70	97.81
[104]	80.09	96.05	89.27	95.80
[154]	91.89	97.10	**98.73**	98.54
[3]	**96.35**	99.33	93.80	**98.61**

imbalanced dataset, and presence of intersubject heterogeneity among images may enforce a system to learn subject-specific features rather than class-specific features. Hence, these factors make the ALL classification more difficult. The systematic review on recent advancements in deep and machine learning-based detection and classification of acute lymphoblastic leukemia (ALL) [9] provides a comprehensive analysis of various models and algorithms used for leukemia detection. The review highlights the performance of different deep learning models such as ResNet18, MobileNetV2, and ShuffleNet, as well as machine learning-based classification schemes like SVM and logistic regression. Specifically, the review emphasizes the superior performance of a hybrid ALL detection model, which combines the advantages of ResNet18 and MobileNetV2, resulting in the best specificity, precision, accuracy, and F1 Score performances in the ALL-IDB1 dataset. Additionally, the MobileNetV2-SVM framework-based ALL classification scheme suggested by Das et al also demonstrates excellent performance with the best accuracy of 99.39% and the best sensitivity of 100.00%. Furthermore, the review discusses the performance of different models in the ALL-IDB2 dataset and provides insights into the benefits of transfer learning-based feature extraction and machine learning-based classification for leukemia detection. Overall, the systematic review offers valuable information on the advancements in deep and machine learning techniques for leukemia detection and classification, shedding light on the potential applications of these technologies in improving the accuracy and efficiency of leukemia diagnosis and treatment. [10] The systematic review identified 17 relevant articles from credible databases, showcasing the expanding use of machine learning methods in peripheral blood smear (PBS) image analysis over the past five years. The average accuracy of machine learning methods applied in PBS image analysis to detect leukemia was found to be over 97%, indicating the extraordinary potential of machine learning in leukemia detection. Deep learning (DL) techniques, in particular, achieved higher precision and sensitivity in detecting different cases of leukemia compared to previous methods. The review also emphasized the significant attention given to the use of machine learning algorithms in detecting acute lymphoblastic leukemia (ALL), suggesting a promising application of machine learning in the field of haematology and artificial intelligence.

III. Methodology

This section presents the details of dataset collection we used, then data augmentation using image transformations and finally applying YOLOV5 and YOLOV7 and YOLOV5 with Senet mechanism to classify the leukemia and healthy cells.

1. Dataset

In this study, we collected data from single source: ALL-IDB dataset. It provides annotated microscopic blood cell images designed for evaluation of segmentation and classification purposes. It included only ALL type of leukemia and HEALTHY samples. The other subtypes of Leukemia (i.e., AML, CML, and CLL) did not exist in this dataset.

The data set consists a total of 260 images. It consists of total 2 target classes: Leukemia, Healthy .

2. Data Preprocessing

i. Auto-Orient: Auto-Orient in preprocessing automatically aligns images to the most suitable orientation, streamlining data consistency and optimizing input for further analysis or processing tasks.
ii. Static Crop: Static Crop in preprocessing involves removing specific, predetermined areas from images, enhancing focus on relevant content and standardizing the visual composition, thereby refining input data for subsequent analysis or applications.
iii. Resize: Images has been resized to 640x640 this standardizes dimensions, facilitating uniformity for downstream tasks such as model training or analysis, while efficiently managing computational resources and ensuring compatibility with specified input requirements.

3. Data Augmentation

Data augmentation techniques were widely utilized to increase the dataset size and avoid memorization, especially in deep learning algorithms. Several image transformation techniques, such as shifting, rotation, and flipping, were employed to obtain different versions of original images. When ML models were trained not only with the original image but also with different image versions, they would have more generalization capabilities. Different studies on image classification with CNNs, various data augmentation techniques reduced the error rate of the model by providing better generalization [11,12,13,14,15,16].

Class	No. of Images
Leukemia cell	130
Healthy cell	130

Table 3: No. of samples of each class in dataset.

In this study, utilized dataset provided a good collection of microscopic blood cell images, however, it contained very limited number (i.e., a couple of hundreds) of samples for each class. The exact number of samples for each class are shown in Table 3. The number of samples was 260 in the original ALL-IDB. Data sample sizes were quite few for deep learning methods, hence it could lead to memorization of the algorithm. Therefore, we increased the number of samples using the following seven image transformations:

i. *Rotation (-45° to +45°)*: It was done by giving the rotating effect to the image in a random direction (left or right). During this process, pixel values of an image are moved left, right, up, down according to the degree value specified between 0–180 as shown in **Figure 2**b. In this study, the degree value was selected at 45° to obtain various images.

ii. *Rotation (90°)*: It was obtained by rotating images by 90 degrees in clockwise, counter- clockwise, upside Down With the help of this transformation, machine learning models avoid memorizing images that were constantly centred in the dataset.

iii. *Saturation (-24% to +24%)*: Applying a 24% augmentation saturation to an image enhances its color vibrancy, striking a balance between vividness and natural tones, optimizing visual appeal for image processing tasks.

iv. *Brightness (-35°% to +35%)*: Implementing a 35% augmentation in brightness for image processing introduces a subtle radiance, illuminating the image without compromising its overall contrast, ideal for achieving an enhanced visual experience

v. *Horizontal Flip*: In this operation, image pixels were moved horizontally from one half of the image to the other half. It was determined that the pixel values would move in a random direction.

vi. *Vertical Flip*: The image was divided by a line drawn horizontally from the centre of the image. As observed pixels were moved vertically unlike horizontal flip.

vii. *Shearing (-15° to +15°)*: It was done by shifting the image pixels counter-clockwise according to the specified angle in degree. In this study, this value set at +15°.

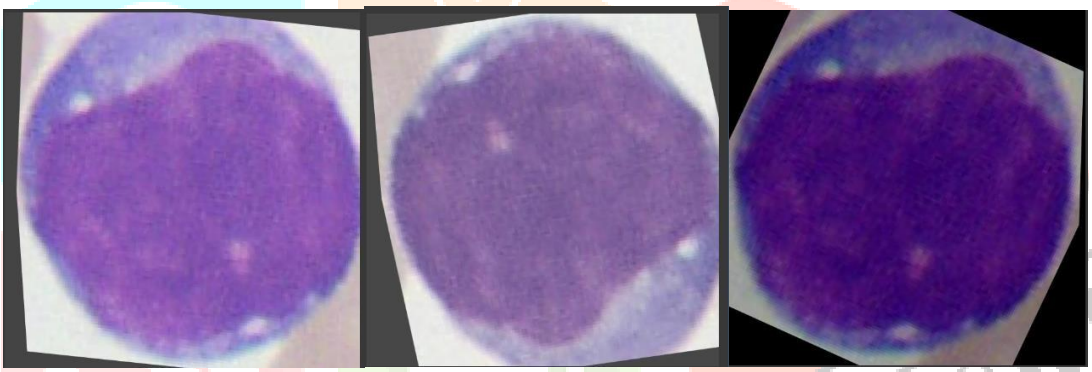


Figure 2. The effect of applying image transformation on one image sample. **(a)** Original image **(b)** shearing, saturation (+13), vertical flip, rotation (15) **(c)** brightness (-), shearing, rotation (90), brightness (-17), saturation (-20)

4. Dataset Split

After the preprocessing we have a total of 624 images. The common split for machine learning tasks is often 70% for training, 20% for validation, and 10% for testing. This allocation ensures that the model learns from a substantial portion of the data, validates its performance on an independent set, and finally tests its generalization on unseen examples, contributing to a robust and reliable model evaluation. But here we have considered a split of 88%, 8% and 4% for training, validation and test respectively so that we have sufficient data to train the model.

- Training Set: 88% of 624 = 546 images
- Validation Set: 8% of 624 = 52 images
- Testing Set: 4% of 624 = 26 images

This partitioning helps in training a machine learning model on a majority of the data, tuning and validating its performance on a separate set, and finally, assessing its generalization on an independent testing set.

5. Network architecture of enhanced YOLOv5

YOLOv5, or You Only Look Once version 5, is an object detection model renowned for its speed, accuracy, and streamlined architecture. It employs a single neural network to predict bounding boxes and class probabilities directly from full images in real-time. YOLOv5 features a CSPDarknet53 backbone, which enhances feature extraction efficiency, and utilizes PANet for improved spatial information integration as shown in Figure 3. One key factor contributing to YOLOv5's superiority lies in its balance between precision and speed. The model maintains high accuracy in object detection while achieving real-time performance, making it suitable for a diverse range of applications, from surveillance to autonomous vehicles. Additionally, YOLOv5 benefits from a simplified architecture, making it more accessible and easier to deploy compared to some complex counterparts. However, it's important to note that the effectiveness of a model often depends on the specific requirements and characteristics of the task at hand. Different models may excel in different scenarios, and the choice of the "best" model depends on factors such as speed, accuracy, and computational

resources available for deployment. In order to get better results on the leukemia dataset which is very different than the COCO dataset which has only 80 classes.

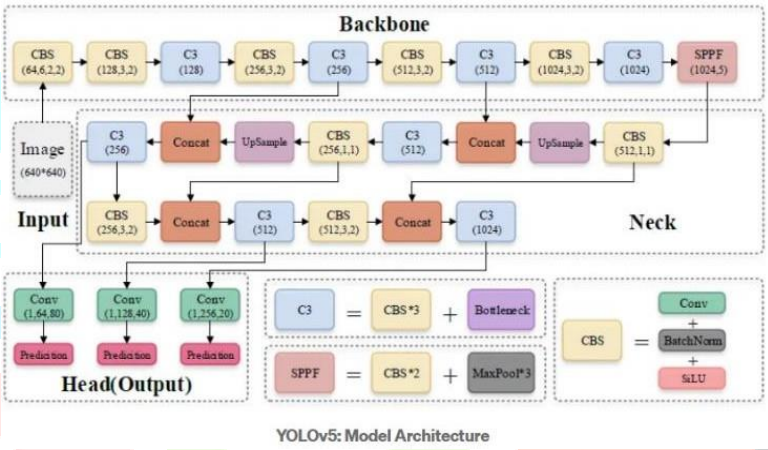


Figure 3. YOLO v5 Model architecture

3.5.1 SENet Attention Mechanism

The attention mechanism [17] is commonly used in data processing, particularly when it's necessary to highlight important features. In cases where smaller features might be overlooked, introducing an attention mechanism like channel attention [18] can help. It works by identifying and amplifying the most relevant channels while reducing the impact of less useful ones, thus helping to keep track of features that might otherwise be lost during feature extraction.

A specific type of attention mechanism known as the Squeeze-and-Excitation Network (SENet) [19], developed by Momenta and Jie H. from the University of Oxford, is designed to increase the representational capacity of a network by modeling the relationships between convolutional feature channels. The architecture of SENet, as shown in Figure 4, uses two main processes: "squeeze" and "excitation."

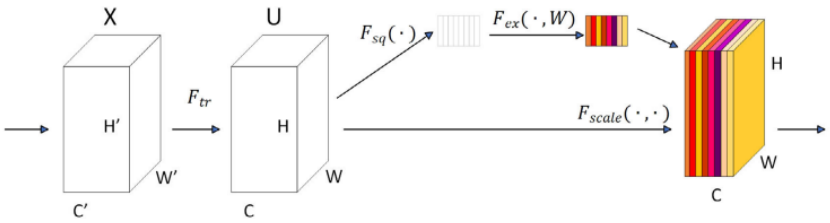


Figure 4. SENet module Architecture

The "squeeze" step compresses spatial information by aggregating features across the height and width dimensions, resulting in a channel descriptor that condenses H × W × C dimensions into 1 × 1 × C, as detailed in Equation 1.

$$Z_c = F_{sq}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \qquad (1)$$

In the "excitation" step, the system learns to adaptively boost certain channels while reducing the influence of others, using global information to guide these decisions. This is achieved through a gating mechanism that utilizes a bottleneck with two fully connected layers, one for reducing dimensionality (W1) and another for expanding it (W2), as shown in Equation 2.

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \qquad (2)$$

However, adding too many modules can reduce computational efficiency. To effectively capture key features from smaller targets without overwhelming the system, this paper incorporates SENet selectively, at specific locations where critical feature layers need to be extracted first each convolution layer.

IV. Experimentation

We created an experimental environment convenient with the designed CNN architecture. We ran all the experiments in a computer with Core i7 processor and 8 GB RAM running under Windows 10 operating system, and Anaconda 3 with Spider 3.3 and Python 3.7. After setting up the environment, we conducted 2 experiments as follows:

- *Experiment #1*: The primary experiment conducts measuring the first capabilities of the YOLOV5 model. It was first trained on COCO dataset for the initialising of the primary weights for the model and then later trained on the Acute lymphoblastic leukemia dataset with the data partitioned for training. The data set consisted of all subtypes of ALL with 546 training, 52 validation and 26 testing images.

- *Experiment #2*: In this experiment we measured the capabilities of the YOLOV7 model. It was first trained on COCO dataset for the initialising of the primary weights for the model and then later trained on the Acute lymphoblastic leukemia dataset with the data partitioned for training. The data set consisted of all subtypes of ALL with 546 training, 52 validation and 26 testing images.

- *Experiment #3*: In this experiment we measured the capabilities of YOLOv5 enhanced with SENet attention module. It was first trained on COCO dataset for the initialising of the primary weights for the model and then later trained on the Acute lymphoblastic leukemia dataset with the data partitioned for training. The data set consisted of all subtypes of ALL with 546 training, 52 validation and 26 testing images.

Furthermore, In the experimental setup, hyperparameters were meticulously tuned for optimal performance in two distinct object detection models, YOLOv5 and YOLOv7 and enhanced YOLOv5. For batch size, epoch, and learning rate, a systematic exploration was conducted to identify the configurations yielding the best results.

For YOLOv5 and YOLOv7, the batch size, representing the number of input samples processed in each iteration, and the number of training epochs, defining the complete passes through the dataset, were fine-tuned to strike a balance between model convergence and computational efficiency. Concurrently, the learning rate, governing the size of parameter updates during optimization, underwent careful adjustment to enhance model training stability and convergence speed.

Similarly, in the case of enhanced YOLOv5, the same hyperparameters - batch size, epoch, and learning rate - underwent a tailored optimization process. This involved iterative adjustments to find the optimal

combination that maximized the model's ability to accurately detect objects while mitigating overfitting or underfitting issues.

The experimental setup aimed to systematically explore the hyperparameter space, considering the unique architectural nuances of YOLOv5 YOLOv7 and enhanced YOLOv5, ultimately achieving peak performance in terms of object detection accuracy and efficiency for each respective model. The final configurations were selected based on their ability to generalize well to unseen data, demonstrating the effectiveness of the carefully tuned hyperparameters in optimizing model performance.

V. Results

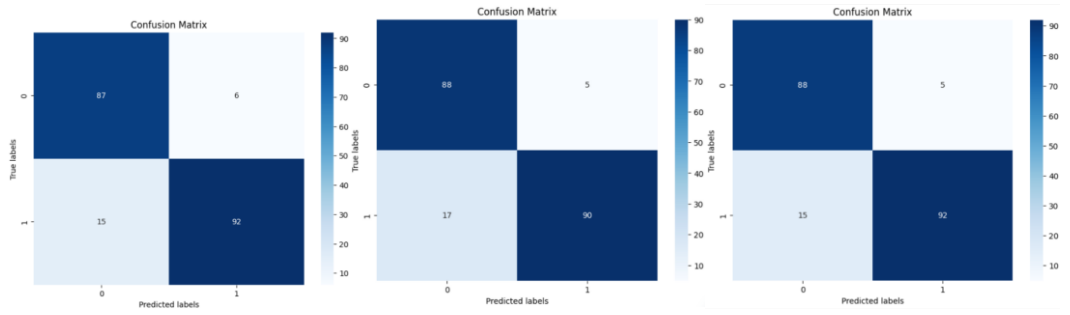
The already performed experimental results for leukemia white blood cell classification are summarized for two-stage classifiers, namely SVM, RandomForest, Decision Tree, Naive Bayes, and VGG-16. It's important to note that these traditional classifiers operate in two stages, unlike YOLO, which functions as a single-stage classifier.

The performance outcomes highlight the efficacy of each method in distinguishing between leukemia white blood cells. SVM, RandomForest, and Decision Tree models demonstrate their classification prowess, leveraging a two-stage approach. Naive Bayes, known for its probabilistic framework, also contributes to the classification task, while the deep learning model VGG-16 exhibits its capability as a two-stage classifier in this context.

In contrast, YOLO stands out as a single-stage classifier, streamlining the process with its real- time object detection capabilities. While the two-stage classifiers showcase their effectiveness, YOLO's unique architecture and single-stage operation underscore its potential for efficient and rapid leukemia white blood cell classification, providing an alternative perspective in the realm of medical image analysis.

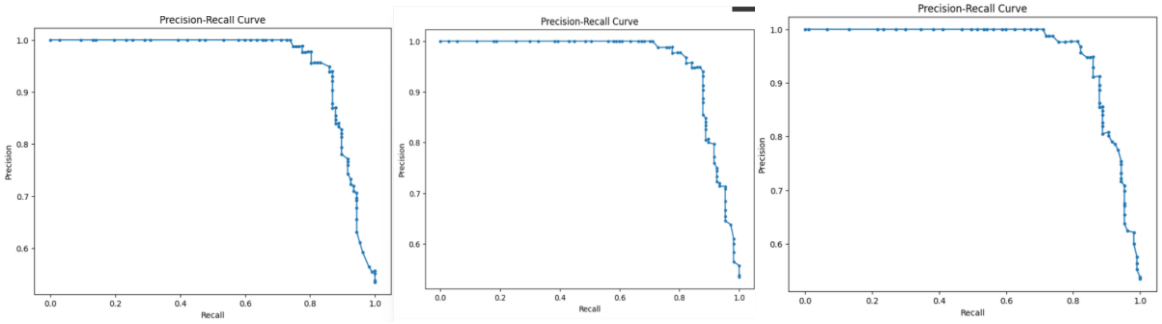
S.No.	Algorithm	Accuracy	Precision	Recall	Specificity	F1 Score
1	SVM	73.02	89.47	53.12	65.9	66.66
2	Random Forest	96.83	100	93.75	93.93	96.77
3	Decision Trees	96.77	94.11	100	100	96.96
4	Naive Bayes	74.6	69.05	90.65	85.71	78.37
5	VGG-16	90.1	84.88	93.58	87.5	89.01

Figure 5. Comparison of 2 Stage Classifiers



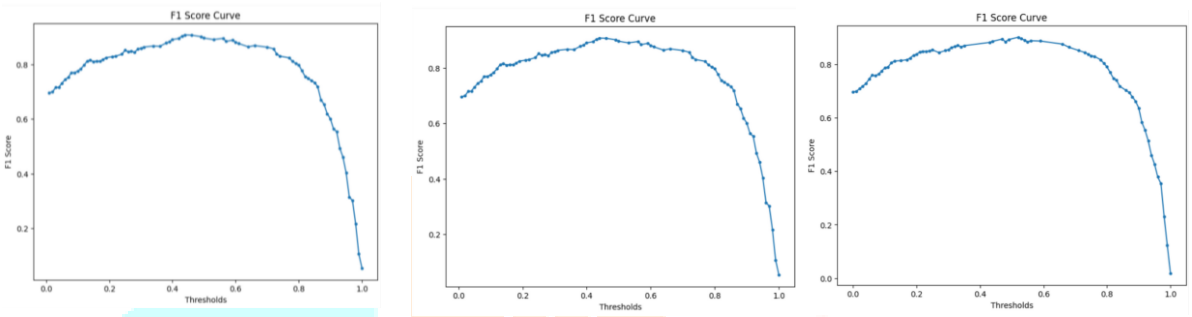
(a) (b) (c)

Figure 6. Confusion matrix of (a) YOLOv5 (b) YOLO v7 (c) Enhanced YOLOv5



(a) (b) (c)

Figure 7. Precision-Recall curve of (a) YOLOv5 (b) YOLOv7 (c) Enhanced YOLOv5



(a) (b) (c)

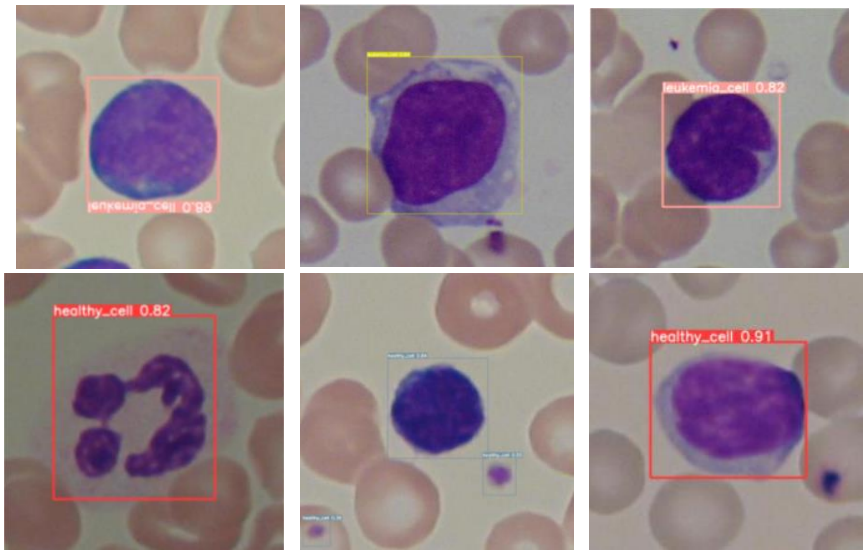
Figure 8. F1-Score curve of (a) YOLOv5 (b)YOLOv7 (c) Enhanced YOLOv5

Performance Comparison:

Among the three models, the model that incorporates the SENet module into YOLOv5 achieved the highest accuracy at 90%. The standard YOLOv5 model showed a slightly lower accuracy at 89.5%, while YOLOv7 had the lowest accuracy at 89%. The inclusion of the SENet (Squeeze-and-Excitation Network) in YOLOv5 seems to have had a positive impact on accuracy, indicating that this architectural enhancement contributes to improved model performance. Despite being the most recent iteration, YOLOv7 showed a slightly lower accuracy compared to YOLOv5 with and without SENet. This might suggest that the improvements in YOLOv7 didn't necessarily translate into higher accuracy in this specific use case.

Model Selection for Application:

Depending on the application's needs and other factors like speed, computational efficiency, or inference time, you would choose the appropriate model. If accuracy is the primary criterion, then YOLOv5 with SENet would be the best choice. However, if other factors are critical, you might consider other aspects beyond accuracy.



(a) (b) (c)

Figure 9. Detection made by (a) YOLOv5 (b)YOLOv7 (c) Enhanced YOLOv5

VI. Conclusions

Leukemia, an aggressive cancer impacting white blood cells and bone marrow, significantly compromises the body's immune system. A prevalent diagnostic approach involves analyzing microscopic blood cell images, specifically blood smears. This study presents a comparative analysis of leukemia diagnosis from microscopic blood images, utilizing different YOLO architectures capable of identifying acute lymphoblastic leukemia subtypes. Our model demonstrates competence in handling a limited number of image samples by incorporating data augmentation techniques, effectively addressing the overfitting issue. Consequently, our enhanced YOLOv5 exhibits superior performance compared to other machine learning algorithms, achieving a remarkable 90% accuracy in the binary classification of one leukemia type (ALL and healthy samples) followed by YOLOv5 with an 89.5% accuracy in classifying leukemia. While medical image classification execution time remains a challenge, it is crucial to assess the model's stability throughout the entire classification process.

In future work, we plan to expand our experiments by using a hybrid deep learning approach using Mask RCNN along with YOLOv5 to enhance the performance. Furthermore, we plan to enlarge our dataset by adding new samples as well as using new data augmentation techniques.

VII. References

[1] D. T. O. Yeung, M. P. Osborn, and D. L. White, ‘‘B-cell acute lymphoblastic leukaemia: Recent discoveries in molecular pathology, their prognostic significance, and a review of the current classification,’’ Brit. J. Haematology, vol. 197, no. 1, pp. 13–27, Apr. 2022.

[2] American Cancer Society. (2022). Leukemia Blood Cancer. [Online]. Available: https://www.cancer.org/cancer/acute-lymphocytic

[3] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu and J. Yao, "DeepPap: Deep convolutional networks for cervical cell classification", *IEEE J. Biomed. Health Informat.*, vol. 21, no. 6, pp. 1633-1643, Nov. 2017.

[4] P. K. Das and S. Meher, ‘‘An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia,’’ Expert Syst. Appl., vol. 183, Nov. 2021, Art. no. 115311.

[5] J. Zhao, M. Zhang, Z. Zhou, J. Chu and F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks", *Med. Biol. Eng. Comput.*, vol. 55, no. 8, pp. 1287-1301, Aug. 2017

[6] R. Gupta, S. Gehlot, and A. Gupta, ‘‘C-NMC: B-lineage acute lymphoblastic leukaemia: A blood cancer dataset,’’ Med. Eng. Phys., vol. 103, May 2022, Art. no. 103793.

[7] S. Goswami, S. Mehta, D. Sahrawat, A. Gupta, and R. Gupta, ‘‘Heterogeneity loss to handle intersubject and intrasubject variability in cancer,’’ 2020, arXiv:2003.03295.

[8] A. Gupta and R. Gupta, C-NMC Challenge: Classification in Cancer Cell Imaging, vol. 10. Singapore: Springer, 2019, pp. 978–981.

[9] Pradeep Kumar Das, Diya V A, Sukadev Meher Rutuparna Panda, and Ajith Abraham (June 2022) A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukaemia

[10] Mustafa Ghaderzadeh , 1 Farkhondeh Asadi , 1 Azamossadat Hosseini , 1 Davood Bashash , 2 Hassan Abolghasemi , 3 and Arash Roshanpour (2021),” Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review”

[11] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Stateline, NV, USA, 3–6 December 2012; pp. 1097–1105. [**Google Scholar**]

[12] Ciresan, D.; Meier, U.; Masci, J.; Gambardella, L.; Schmidhuber, J. High-performance neural networks for visual object classification. *arXiv* **2011**, arXiv:1102.0183. [**Google Scholar**]

[13] Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the International Interdisciplinary PhD Workshop (IIPhDW), Swinoujście, Poland, 9–12 May 2018; pp. 117–122. [**Google Scholar**]

[14] Wong, S.C.; Gatt, A.; Stamatescu, V.; McDonnell, M.D. Understanding data augmentation for classification: When to warp? In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 30 November–2 December 2016.

[15] Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621. [**Google Scholar**]

[16] Shijie, J.; Ping, W.; Peiyi, J.; Siping, H. Research on data augmentation for image classification based on convolution neural networks. In Proceedings of the Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 4165–4170. [**Google Scholar**]

[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ‘‘Attention is all you need,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 1–11.

[18] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, ‘‘Attention mechanisms in computer vision: A survey,’’ Comput. Vis. Media, vol. 8, pp. 331–368, Mar. 2022.

[19] J. Hu, L. Shen, and G. Sun, ‘‘Squeeze-and-excitation networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.