Research article

Diagnosis of Blood Microscopic Sample for Early Detection of Acute Lymphoblastic Leukemia Using Computer Vision Based on The Object Detection Model YOLO V5.

By: Donia Mohamed, Perryhan Essam, Maha Emad, Gehad Essam, Shehab Rashad

Department of Medical Informatics, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

Abstract

Acute lymphoblastic leukemia (ALL), a disease that is quite common, necessitates invasive, costly, and time consuming diagnostic procedures for a final diagnosis. The early classification of cancer cases from noncancerous samples plays a crucial role in ALL detection utilizing peripheral blood smear (PBS) images. Due to the general character of all signs and symptoms, misdiagnosis is a common concern when these PBS images are analyzed by laboratory. This research highlights the successful application of the YOLOv5 object detection model in detecting ALL cells, achieving 97.8%, 96%, 96.7%, accuracy, precision, and recall respectively using an ALL image dataset after adding the bounding boxes manually. The findings demonstrate the potential for developing computer-aided diagnostic systems that augment healthcare professionals' capabilities and improve patient outcomes through early and accurate detection of ALL.

Keywords: Acute lymphoblastic leukemia (ALL); YOLO V5; Computer vision; Object detection

1. Introduction

ALL is a malignancy of B or T-lymphoblasts characterized by uncontrolled proliferation of abnormal, immature lymphocytes and their progenitors which ultimately leads to the replacement of bone marrow elements and other lymphoid organs resulting in a characteristic disease pattern [1]. To illustrate, Health issues including leukemia, thalassemia, and anemia can result from an increase or reduction in any of the fundamental blood components. A high WBC volume decreases body immunity because it covers both platelets and RBCs. ALL, chronic lymphocytic leukemia, acute myeloblastic leukemia, and chronic myeloblastic leukemia are the four categories according to their development, pace, and effects. The most prevalent and lethal of them is ALL, which accounts for 70% of all instances of leukemia. The progression of the illness is also significantly influenced by environmental and genetic variables. ALL is brought on by the bone marrow's excessive and unchecked multiplication of lymphocytes [2]. With about 6500 cases annually in the United States alone, ALL is the second most prevalent acute leukemia in adults [3]. ALL is becoming more common within the healthcare system every year, with 64.2 thousand new cases of ALL diagnosed globally between 1990 and 2017. The most common form of leukemia in kids is ALL, which also has a high prevalence in adults [4]. The difficulty in making an early diagnosis of lymphocytes stems from the similarities between normal and lymphoid cell types. As a result, lymphocytes were divided into three groups: reactive, atypical, and normal. The characteristics of normal lymphocytes include homogeneity and round, tiny, and rough nuclei; those of atypical cells include a big size and nucleus as well as the presence of lumpy chromatins; and those of reactive cells include heterogeneity and the presence of red cells around them. The lymphocyte types are identified by microscopic examination, which requires the collection of

blood or bone marrow samples for a pathologist to analyse [5]. However, obtaining a sample of bone marrow and doing an analysis on it are necessary for a proper leukemia diagnosis. The analysis is laborious, time-consuming, and sensitive to the various expert viewpoints because it is done manually. Consequently, notwithstanding the possibility of human mistake, an accurate manual diagnosis depends on the pathologist's ability. In order to automatically detect leukemia, some researchers have suggested collecting WBC characteristics from microscopic pictures. Therefore, the quick and accurate diagnosis made possible by the automatic recognition of blood cell pictures will allow for the evaluation of several cells from each individual. Manual diagnostic issues can be resolved by machine and deep learning approaches. It has been demonstrated that the YOLOv5 which has a greater capacity to distinguish between normal and blast cells, can assess and address many of the issues associated with manual diagnosis and medical imaging [6].

2. Related work

Numerous researches were conducted for implementing deep learning models techniques to classify various types of cancers. However, a recent study proposed GBHSV-Leuk method that highlights the importance of the image pre-processing stage for improved segmentation and classification performance of the algorithm. The HSV color space is used in the proposed method for the segmentation technique to extract the nucleus, as it yields better results in contrast to RGB color space. The experiment was carried out on a hybrid dataset comprising both a real-time dataset and the publicly available ALL-IDB1 dataset. The comparison results show that the proposed method gives an improved result, with the accuracy obtained with the private dataset being 96.30% and the accuracy obtained with the public dataset being 95.41% [7]. Different which is about white blood cell (WBC) detection and counting on Microscopic Blood Cell images automatically can support expertise in diagnosing Acute Lymphoblastic Leukemia (ALL) more easily and quickly. It was about comparing the performance of the YOLO and Mask R-CNN models for the detection of ALL subtypes using evaluation metrics such as precision, recall, F1, and mAP. The results show that the YOLOv4 outperforms YOLOv5 and Mask R-CNN in the detection of ALL subtypes. The YOLOv4 model has slightly better performance than the YOLOv5 model and Mask R-CNN, with an F1 value of 89.5% and a mAP value of 93.2%, respectively [8]. Based on the ALL-IDB dataset of microscopic blood smears, the utilization of the YOLOv5 model is proposed for real-time detection of leukemic cells. Specifically, the recommendation is made for the use of YOLOv5s, as the smaller model performs inference with 97.2% mAP accuracy, almost twice as fast as YOLOv5m, without sacrificing much accuracy. It is suggested by the results that object detection models can offer more detailed, efficient, and practical alternatives to traditional image classification. Additionally, the implementation of YOLO models for image level classification can be considered if such information is deemed necessary [9]. In another paper, the performance of the YOLO and Mask R-CNN models for the detection of ALL subtypes using evaluation metrics such as precision, recall, F1, and mAP is thoroughly compared. The obtained results clearly indicate that the detection of ALL subtypes is outperformed by the YOLOv4 model when compared to YOLOv5 and Mask R-CNN. It is worth noting that the YOLOv4 model exhibits slightly superior performance in comparison to both the YOLOv5 model and Mask R-CNN, with an F1 value of 89.5% and an mAP value of 93.2%, respectively. The object detection and instance segmentation techniques employed in this study offer significant advantages, as they require only a single learning framework, eliminating the need for various separate stages such as WBC segmentation, touch cell separation, feature extraction, and classification. By utilizing these techniques, the process of supporting expertise in diagnosing Acute Lymphoblastic Leukemia (ALL) becomes easier and quicker. Notably, the YOLOv4 model's superior performance can be attributed to its advanced architecture and features, which enable robust object detection and precise instance segmentation. The model's ability to efficiently and accurately detect and classify WBCs contributes to its overall success in diagnosing ALL subtypes. As a result of this study, it is evident that the automated detection and counting of White Blood Cells (WBCs) in microscopic blood cell images greatly aids in streamlining and expediting the diagnostic process for ALL subtypes. By leveraging the YOLOv4 model's remarkable performance, medical professionals can benefit from a more efficient and reliable tool for identifying and localizing ALL subtypes within the images[10].

3. Methods

This section displays the methods used to analyze ALL image dataset for early detection of lymphoblastic leukemia. The dataset has samples from people with different stages of lymphoblastic leukemia ALL; Benign, Early, Pre, Pro. All images went through noise reduction, and various types of augmentation techniques then a YoloV5 model was used for ALL types sub-classification.

3.1 Dataset description

This study evaluates object detection model to be specific on ALL image dataset. The bone marrow laboratory at Taleqani Hospital provided the images for this dataset (Tehran, Iran). This dataset included 3256 PBS pictures from 89 individuals who were thought to be ALL and whose blood samples were properly processed and stained by professional laboratory personnel. This dataset is separated into the benign and malignant classes. Hematogenes make up the first group, whereas the ALL group, which includes the three malignant lymphoblast subtypes Early Pre-B, Pre-B, and Pro-B ALL, makes up the latter. All of the photos were taken with a Zeiss camera at a 100x magnification in a microscope, and they were all saved as JPG files. The specific types and subtypes of these cells were identified by a professional using the flow cytometry instrument. Figure 1 illustrates some samples from ALL image dataset [11].

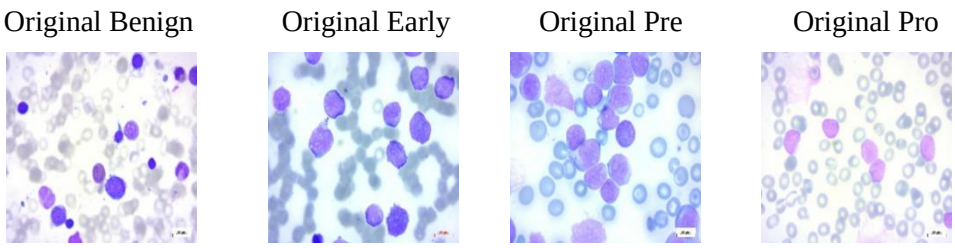


Figure 1. Samples from (ALL) image dataset

3.2 Object detection and YOLO v5:

3.2.1 Object detection

Object detection is a computer vision task that involves identifying and localizing specific objects within an image or a video. The goal of object detection is to not only determine the presence of objects in an image but also to accurately delineate their boundaries or regions of interest. It goes beyond image classification, which focuses on assigning a single label to an entire image, by providing a more detailed understanding of .the visual content

Object detection algorithms typically work by analyzing the entire image or video frame and identifying the objects of interest, usually by predicting their bounding boxes and assigning class labels to them. These algorithms leverage machine learning techniques, particularly deep learning, to learn patterns and features .that are indicative of different object categories

There are various approaches to object detection, including region-based methods, sliding window methods, and single-shot methods. Region-based methods, such as the Faster R-CNN (Region-based Convolutional Neural Network), propose regions of interest within the image and then classify and refine these regions to generate accurate object detections. Sliding window methods involve scanning the image with a fixed-size window at different scales and positions to identify potential objects. Single-shot methods, such as YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector), directly predict object bounding boxes .and class labels in a single pass over the image

Object detection can be found in many applications in various fields, including autonomous driving, video surveillance, robotics, augmented reality, and most importantly medical imaging. It enables machines to

understand and interact with their visual environments, opening up possibilities for automated analysis, object tracking, and intelligent decision-making based on visual cues[12].

Bounding boxes *were set manually by the team* so the model can be trained using the ALL dataset via Roboflow. Here are some samples of images after adding the bounding boxes for object detection.

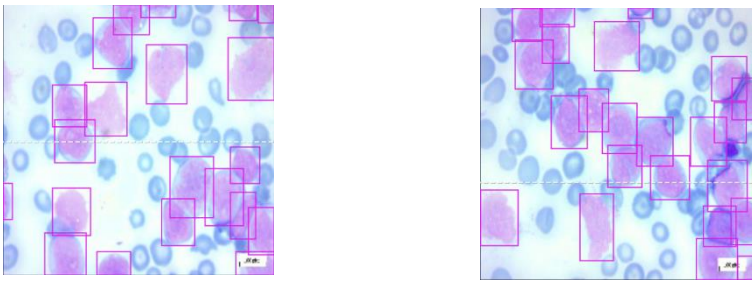


Figure 2. Samples from (ALL) image dataset after adding the bounding boxes manually

3.2.2 YOLOv5

YOLO (You Only Look Once) is a popular and influential object detection algorithm that has seen several iterations in its development. YOLO v5 represents the latest iteration of this algorithm, offering significant improvements and advancements over its predecessors. It has gained attention and widespread adoption in the computer vision community due to its exceptional speed and accuracy in detecting objects within images and videos. The fundamental idea behind YOLO is to treat object detection as a regression problem, where a single neural network predicts bounding boxes and class probabilities directly from the input image in a single pass. This differs from other approaches that rely on region proposal techniques, which are computationally expensive and time-consuming. YOLO v5 builds upon the strengths of its predecessors while introducing notable improvements. It incorporates a more advanced architecture, leveraging state-of-the-art deep learning techniques and advancements in neural network design. The architecture comprises a backbone network for feature extraction, followed by multiple detection heads that process different scales and resolutions of the feature maps. This multi-scale approach allows YOLO v5 to capture objects of varying sizes with improved accuracy. Moreover, YOLO v5 introduces a streamlined and efficient design, making it faster and more lightweight compared to previous versions. The model achieves a remarkable balance between speed and accuracy, enabling real-time object detection even on resource-constrained devices. This efficiency is achieved through the use of various optimization techniques, such as network slimming, model pruning, and network architecture enhancements. The training process of YOLO v5 involves collecting a large annotated dataset and utilizing advanced data augmentation techniques to improve the model's generalization capabilities. The model is then trained using state-of-the-art optimization algorithms, enabling it to learn robust representations of objects and their respective classes.

The applications of YOLO v5 are diverse and span across various domains. It has proven to be highly effective in object detection tasks, including but not limited to pedestrian detection, vehicle detection, object tracking, and medical image analysis. The model's speed, accuracy, and efficiency make it suitable for real-time applications, video analysis, and scenarios where fast and accurate object detection is crucial.

3.3 Image Enhancement

Image preprocessing and augmentation are two of the most powerful tools that if used correctly it can enhance the image quality hence the results to great extent [7]. Therefore, for original dataset some preprocessing techniques were applied for better accuracy. Zoom 10%, rotation 10, brightness was set randomly from 0.8 to 1.5, and vertical flip and horizontal flip were applied. Moreover, images width and height were set to 224 and rescaled to 1/255.0, Table 1 displays the augmentation and preprocessing used.

Table 1. Image preprocessing and augmentation techniques used

Technique used	Range/Type
Zoom range	10%
Rotation range	10
Brightness	0.8 – 1.5
Rescale	1/255.0
Horizontal flip	True
Vertical flip	True

3.1.4 Training

As mentioned previously, we chose to use segmented cells to train the network to lymphoblast cells, improve the model's accuracy and reliability, and provide a more precise attention map after building and testing various fitting structures and techniques. However, we noticed that much essential information is lost when segmented cells are used instead of the original image. Hence, we considered segmented cells and the original images as two inputs for the network. Each input sample was an image pair that included both the segmented and PBS versions of the same image. Segmented images and original images were divided into training set, testing set, and validation set as shown in Table 2.

Table 2. How both data from original and segmented datasets were divided between sets

	Train set	Test set	Validation set	Total
Percentage (%)	64	20	16	100
No. of paired samples	2083	652	521	3256

4. Results

The effectiveness of the suggested method is assessed through different techniques and the findings are presented in this section.

4.1 Parameter tuning

All the parameters started operating with random weights for network training. The batch size and number of epochs in the suggested technique were set to 100 and 32, respectively. The classes were weighted proportionately to the number of samples in order to address the issue of unbalanced data. The use of the Adam optimizer with a 1e 3 starting learning rate helped the gradient descent process perform better. Table 3 displays the techniques used for parameter tuning.

Table 3. Techniques used for parameter tuning

	Batch size	Epochs	Learning rate
Value	32	100	0.001

4.2 Evaluation metrics

In this study, clinically significant statistical parameters like precision, accuracy, and precision are used to assess the performance of the proposed detection and classification model. Unfortunately, precision is not always sufficient to assess the model's performance, particularly when dealing with an asymmetrical data set. As a result, different performance measures must be assessed in order to test the model [13].

1- Mean Average Precision (MAP) is a widely used evaluation metric in information retrieval and machine learning tasks, particularly for tasks involving ranked retrieval systems, such as document retrieval, image retrieval, and object detection. Furthermore, mean Average Precision extends the evaluation to multiple retrieval cutoffs. It computes the average precision (AP) for each query or instance in a dataset and then takes the mean of these average precisions. AP is calculated by taking the precision values at various cutoff points (e.g., ranks) and averaging them by considering the ground truth relevance of the retrieved items. Equation1 and equation2 show the AP and MAP equations, respectively.

Equation1. Average precision

$$AP = \frac{\sum P}{Num(objects)}$$

Equation2. Mean average precision

$$mAP = \frac{\sum AP}{Num(class)}$$

2- Precision: In deep learning, precision is a commonly used evaluation metric to measure the performance of a classification model. It is a measure of the accuracy of the positive predictions made by the model. Precision is calculated as the ratio of true positives (TP) to the sum of true positives and false positives (FP). Equation3 shows how precision is calculated.

Equation3. Precision

$$P = \frac{TP}{TP} + FP$$

True positives (TP) are the number of correctly predicted positive samples, while false positives (FP) are the number of negative samples that were incorrectly predicted as positive. Precision specifically focuses on the positive predictions made by the model and provides insights into the model's ability to minimize false positives. A high precision value indicates that the model has a low rate of false positives, meaning that when it predicts a positive outcome, it is likely to be correct. For example, in a binary classification problem where the model is distinguishing between cats and dogs, precision measures the proportion of predicted cat samples that are actually cats, as opposed to incorrectly predicting a dog as a cat. It is important to note that precision is often used in conjunction with other metrics like recall, accuracy, and F1 score to get a comprehensive understanding of the model's performance. Precision is especially useful when the cost of false positives is high, such as in medical diagnosis or spam detection, where false positives can have significant consequences.

3- Recall: In deep learning, recall is another commonly used evaluation metric to measure the performance of a classification model. It is a measure of the model's ability to correctly identify all relevant positive samples. Recall, also known as sensitivity or true positive rate (TPR), is calculated as the ratio of true positives (TP) to the sum of true positives and false negatives (FN). Equation4 shows how recall is calculated.

Equation4. Recall

$$R=\frac{TP}{TP}+FN$$

True positives (TP) are the number of correctly predicted positive samples, while false negatives (FN) are the number of positive samples that were incorrectly predicted as negative. Recall specifically focuses on the positive samples that were correctly identified by the model, regardless of whether there were any false positives. A high recall value indicates that the model has a low rate of false negatives, meaning that it is effectively capturing the positive instances in the dataset. For example, in a binary classification problem where the model is distinguishing between cats and dogs, recall measures the proportion of actual cat samples that were correctly identified as cats, as opposed to incorrectly predicting a cat as a dog. Recall is particularly important when the cost of false negatives is high, such as in medical diagnosis or anomaly detection, where missing a positive sample can have severe consequences. It ensures that the model is capable of detecting as many positive instances as possible.

Similar to precision, recall is often used in conjunction with other metrics like precision, accuracy, and F1 score to gain a comprehensive understanding of the model's performance.

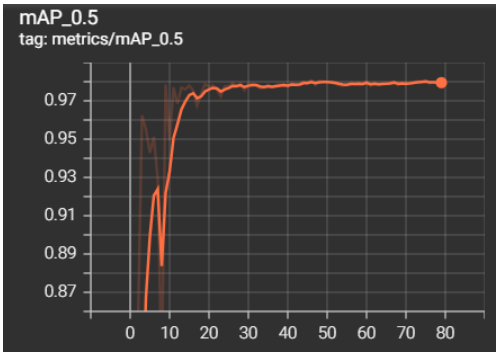
Table 4 shows the results of each evaluation metric used

Table 4. Results of YOLOv5

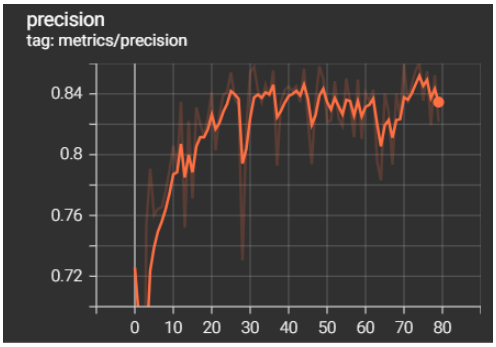
mAP	Recall	Precision
97.8%	96.7%	96%

Figure3. Visual representations for the results

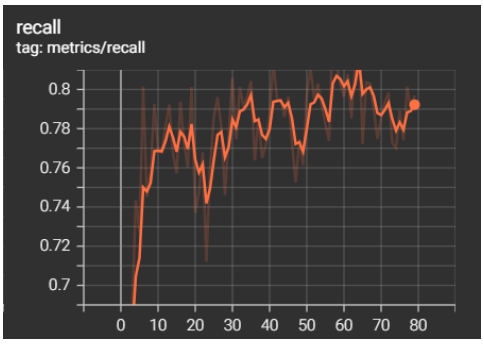
mAP



Precision



Recall



5. Future directions

There are various things that are put into consideration for future enhancement. For instance, it is planned to increase the GPU capabilities from a remote server to implement 220 epoch as this will enhance the performance. It is planned to use more user-friendly GUI. Furthermore, it is planned to use an enhanced encryption algorithm to fully secure the patient’s data.

6. Conclusion

In conclusion, this study has successfully demonstrated the effectiveness of YOLOv5, an object detection model, in accurately detecting Acute Lymphoblastic Leukemia (ALL) and classifying it into four distinct types. The model achieved remarkable classification accuracy with a mAP (mean Average Precision) of 97.8%, recall of 96.7%, and precision of 96%.

The high mAP score reflects the consistent and accurate localization and classification of ALL cells into the four types within the microscopic images. The recall score demonstrates the model's ability to identify the majority of positive cases, ensuring comprehensive coverage in detecting ALL. Additionally, the precision score highlights the model's capability to minimize false positives, resulting in a high degree of accuracy in classifying the different types of ALL cells.

These outstanding results underline the potential of YOLOv5 as a valuable tool for precise and efficient ALL detection, classification, and subsequent treatment planning. The model's high accuracy provides confidence in its ability to assist medical professionals in accurate diagnosis and appropriate patient management.

The remarkable precision and recall scores underscore the model's ability to precisely identify and classify the four types of ALL cells, reducing the risk of misclassification and supporting effective treatment decisions. The accurate identification of specific ALL subtypes aids in tailoring treatment approaches and improving patient outcomes.

While this study presents promising findings, further validation on diverse datasets from different medical institutions is recommended to assess the model's generalizability. Incorporating larger datasets with varied populations and sample variations would enhance the model's robustness and reliability in real-world clinical settings.

7. References

1. Puckett Y, Chan O. ALL. [Updated 2022 Jun 27]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459149/

2. Terwilliger T, Abdul-Hay M. ALL: a comprehensive review and 2017 update. Blood Cancer J. 2017 Jun 30;7(6):e577. doi: 10.1038/bcj.2017.53. PMID: 28665419; PMCID: PMC5520400.

3. Sawyers CL, Denny CT, Witte ON. Leukemia and the disruption of normal hematopoiesis. Cell. 1991 Jan 25;64(2):337-50. doi: 10.1016/0092-8674(91)90643-d. PMID: 1988151.

4. Dong Y, Shi O, Zeng Q, et al. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. Exp Hematol Oncol. 2020;9(1):1-11.

5. Bain BJ. Diagnosis from the blood smear. N Engl J Med. 2005 Aug 4;353(5):498-507. doi: 10.1056/NEJMra043442. PMID: 16079373.

6. Abunadi I, Senan EM. Multi-Method Diagnosis of Blood Microscopic Sample for Early Detection of ALL Based on Deep Learning and Hybrid Techniques. Sensors (Basel). 2022 Feb 19;22(4):1629. doi: 10.3390/s22041629. PMID: 35214531; PMCID: PMC8876170.

7. Devi TG, Patil N, Rai S, Philipose CS. Gaussian Blurring Technique for Detecting and Classifying Acute Lymphoblastic Leukemia Cancer Cells from Microscopic Biopsy Images. *Life*. 2023; 13(2):348. https://doi.org/10.3390/life13020348

8. C. Fatichah, N. Suciati, T. Mustaqim and A. R. Revanda, "Detection of Acute Lymphoblastic Leukemia Subtypes using YOLO and Mask R-CNN," 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2022, pp. 413-418, doi: 10.1109/ISRITI56927.2022.10052976.

9. Emma Chen, Mikhail Y. Shalaginov, Rory Liao, Tingying Helen Zeng
bioRxiv 2022.10.22.513362; doi: https://doi.org/10.1101/2022.10.22.513362

10. C. Fatichah, N. Suciati, T. Mustaqim and A. R. Revanda, "Detection of Acute Lymphoblastic Leukemia Subtypes using YOLO and Mask R-CNN," 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2022, pp. 413-418, doi: 10.1109/ISRITI56927.2022.10052976.

11. https://www.kaggle.com/datasets/mehradaria/leukemia (Accessed 15/5/2023)

12. Horvat, Marko & Jelečević, Ljudevit & Gledec, Gordan. (2022). A comparative study of YOLOv5 models performance for image localization and classification.

13. Shahriar, Md & Li, Huyue. (2020). A Study of Image Pre-processing for Faster Object Recognition.