An Interpretable Yolov8 Framework for Automated Leukemia Detection in Hematology Images

G. Nivethasri
Research Scholar
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research
Chennai, India

Adhithiyan K
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research
Chennai, India

Afraz S
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research
Chennai, India

Balaji R
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research
Chennai, India

Basith M R
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research
Chennai, India

Abstract—Leukemia is a hematological cancer that is life threatening where early and accurate diagnosis is important in enhancing the survival of the patients and in informing them on the right course of action in terms of treatment. Extremity exploration of peripheral bodily fluid vilification Ansehen is strenuous, time-consuming and prone to bury percipient variance, which emphasizes the importance of automated and dependable computer-assisted diagnostic systems. In this survey, our recommendation is a solid acquisition-based method to the automated detection and localization of leukemia cells with the YOLOv8 object detection framework.. The theoretical account was trained and validated with a large dataset of annotate microscopic images of blood smears of various morphological features of leukemic cells. Data attractive, increase, and transfer learning methods were used to improve the model generalization and hardiness. The proposed YOLOv8 showed great detection and high preciseness, high call back, and mean average precision, and was time period inference capable in clinical screening settings. The comparison of the proposed model with the traditional image processing algorithms and the available deep learning strategies showed that the proposed model is characterized by better detection speed and competitive accuracy. According to the results of the inquiry, the projected system can be an effective and reliable instrument in automated leukemia cell detection to lessen the workload of the diagnostic procedure and assist hematologists with the process of large-scale screening and early diagnosis.

Keywords— Leukemia detection; Peripheral blood smear; Medical image analysis; Hematological malignancy.

I. INTRODUCTION

The leukaemia is a heterogeneous cluster of haematological malignancies that are typified by the anarchic proliferation of deviant white blood cells in the bone marrow and marginal blood fluid. It is still among the life-threatening blood cancers on a global scale and it affects both children and adult population and contributes a significant percentage to the proportion of. cancer-related morbidity and mortality. Early and accurate diagnosis is crucial for improving patient survival rates, guiding treatment planning, and reduction the load on health care systems. Traditionally, leucaemia detection relies on microscopic examination of stained peripheral blood smears and bone marrow aspirates by expert hematologists.

This is a manual procedure of recognizing minute morphological changes in leukocytes; nuclear shape, cytoplasmic texture, chromatin pattern and cell size which is very difficult to master, time consuming and subject to inter-observer variability. In areas such as lack of access to trained experts and high patient volumes, the delayed diagnosis and mismatched diagnostic results are are public, light weight the urgent necessity to have automated, reliable, and interpretable computer-aided systems of diagnostic processes.The advances in digital pathology and medical imaging in the recent years enabled the receipt of high-resolution hematology photographs which facilitates the development of new opportunities in the area of unreal intelligence analysis. In the medical image categorization, cleavage and physical object sensing challenge, deep learning, in particular, the convolutional neural network demonstrated performance of incredible results.

The techniques are capable of acquiring complex hierarchical features directly out of raw pixel data, without requiring hand written feature engineering.Object detection frameworks are particularly the most appropriate of the numerous deep learning paradigms in the context of hematology image analysis since it can detect and classify multiple cells in a single microscopic field. Such ability will be crucial to the detection of leukemia, where a critical diagnosis is required concerning the presence, proportion, and morphology of blast cells.YOLO family of models has already become one of the most successful and accurate real-time detectors of objects. YOLO is also fast and does not influence accuracy; casting the detection as a single regression problem, which predicts bounding boxes and class probabilities in a forward pass, the model can be used. The latest, YOLOv8, has been targeted at architecture improvement, such as anchor-free detection, feature pyramid network and training recommendations, especially able to detect small, densely populated objects, such as

blood cells. Its portability can be used in resource-constrained systems, such as point-of-care diagnostic systems and edge devices, which makes it very applicable to low- and middle-income healthcare systems.Although these are the benefits, the majority of the currently available deep learning-based leukemia detection systems are treated as black boxes so that they give predictions without reasoning why this is the case. Interpretability is not a desirable aspect in clinical practice, but a necessity. To rely on the results of a model and prove that it is right, hematologists have to know which morphological features affected the choice of a model to be used in order to include it into the usual working procedures. The absence of transparency may also obstruct regulatory authorization, decrease clinician trust, and limit its use in practice.

Moreover, automated leukemia detection systems have to deal with a number of practical issues that are inherent in the hematology imaging. The images of blood smears are characterised with a high degree of variability with respect to the staining procedures, magnification of the microscope, light, and methods of slide preparation. The cells can either overlap, be partially covered or have morphological similarities in different classes making them hard to detect. Another acute problem is that the number of blast cells can be small in comparison with normal leukocytes. An effective YOLOv8-based system should include strategie that includes information increase, stain normalization, multi graduated table feature extraction, and optimization of loss function as a way of enhancing generalization to a wide range of datasets. Moreover, the annotation with domain experts should be done carefully to obtain high-quality training data because the recognition of an object is directly correlated with the precision of its bounding box..

II. Literature Review

An artificial intelligence-based diagnostic system was suggested by C. S. Vomanesh et al. (2023) [1] to identify leukemia based on the complete blood count (CBC) data and not on images. Their research centered on the training of machine learning algorithms on a priorly preprocessed hematological dataset and their evaluation through the use of accuracy, precision and recall metrics. The paper has cited the benefits of CBC-based models such as low price, quicker processing time, and automation in comparison to the traditional laboratory procedures. Nevertheless, the methodology is weak due to the absence of a morphological cell study, which is needed in order to classify the subtypes and confirm the visualization. Their results form a basis of non-image-based screening on leukemia but insist that the models should be more detailed and include the morphology of the cells.

Q. Fan et al. (2023) [2] proposed the use that a deep convolutional neural network known as QCResNet was used to classify Acute Lymphoblastic Leukemia (ALL) images using peripheral blood smears. Their model, where they used a dataset of 15,135 images, provided an accuracy of 98.9% that is relatively close to some of the existing state-of-the-art methods. It was demonstrated that deep residual learning was achieved in the experiment. is effective to extract discriminative features of hematology images and allow quick diagnosis. Although it is a high-performance model, the model is mainly a black box classifier and lacks localization of the leukemic cells or interpretability which are essential as far as clinical adoption is concerned.

A. Y. et al. (2023) [3] came up with a better CNN-based leukemia detector system to classify microscopic images of blood cells as either leukemic and healthy. Its work also focused on the significance of early diagnosis especially the high rates of leukemia among pediatric patients. The suggested model facilitated and automated a diagnostic process and minimized the reliance on manual microscopy. Nevertheless, the investigation restricted itself to binary classification and multi-class leukemia subtypes were not considered, which suggests that the study is not applicable to actual clinical hematology workflows.

Another study by A. Ramagiri et al. (2023) [4] undertook a comparative study of device acquisition and deep acquisition in leukemia image classification. The authors talked about the problem of manual diagnosis when the leukemic cells are similar to ordinary ones and the increasing role of CNN-based procedures in the decisions making process is important to enhance the accuracy of predictions. Their paper has examined various algorithms to determine the most effective method of leukemia detection. Although the work was insightful on the trend of performance, it was centered on the image-level classification and not cell-level detection and did not have an interpretable framework.

The article by M. Ashok et al. (2023) [5] is an investigative work about automated Acute Lymphoblastic Leukemia detection based on computer vision. They used the method of transforming blood smear images into CMYK color space, segmentation with K-means, and classification with supervised machine learning models, including SVM and XGBoost. It was a hybrid pipeline which proved to be more detective as it targeted nuclear features. Nevertheless, the multi-stage approach is more complex to compute and might not be as efficient as end-to-end deep learning models. Also, deep feature learning is not scaled to large and diverse datasets by the absence of deep feature learning.

The study of Zarish et al. (2022) [6] determined the possibilities of device acquisition algorithms to foretell the survival rates of leukemia patients based on the SEER clinical database. They used multi-class classification hierarchy which classifies patients into intervals of survival and their predictive performance was better. In contrast to the study on diagnostic studies that are based on images, the work tackled the aspect of prognostic modeling and emphasized the need to incorporate clinical variables in leukemia research. The model is not useful as a diagnostic tool but complementary to automated cell detection or morphological analysis despite the fact that it is biologically important.

J. Sheet et al. (2023) [7] suggested a framework of transfer learning that was implemented on MobileNetV2 and CNN models to identify leukemia in microscopic images. Their computer-aided diagnosis system had an accuracy of 96.58 percent indicating the efficiency of lightweight pretrained models in medical image analysis with high sensitivity and specificity.

In one of the studies, K. K. Jha et al. (2022) [8] constructed an ensemble deep learning model to identify leukemia with augmented datasets on Kaggle. Their artificial neural network was able to reach a maximum accuracy of 100 percent on high-quality data and was also able to perform well on low-quality data. The ensemble method enhanced robustness and minimized mean square error which means that the classification can be done with a lot of reliability. Nevertheless, it was not mentioned in the study that real-time detection, cell localization, or interpretability were studied, and the extremely high accuracy could be the evidence of possible overfitting because of the lack of diversity in the datasets.

In [9], R. Arif et al. recommended a multi-stage heavy acquisition model that involved preprocessing, data augmentation, segmentation and classification with the help of a trained CNN and pretrained AlexNet on the ALL-IDB dataset. Their model was very successful in identifying leukemia with an accuracy of 98.05 percent, recall of 100 percent, and large F1-score. The segmentation allowed was

included and increased attention towards pertinent regions of the cell. In spite of these, the pipeline is computationally intensive and does not have a real-time object detection feature as well as visualization of the model decisions.

In [10], the authors A. M. Basymeleh et al. compared various CNN models with transfer learning to analyze leukemia images in multi-class. Their experiments indicated that VGG16 was able to triumph compared to other models and reached an accuracy of 97.50, sensitivity and specificity were high. The paper also demonstrated the data augmentation and the transformation of HSV color space as effective in enhancing the model generalization. Nonetheless, the method lacks cell-level detection, localization, or interpretability, as does other classification-based methods, so this methodology has limited clinical application.

III. METHODOLOGY

The projected system will automatically examine the malignant images of microscopic fringe body fluids with the use of the YOLOv8 object detection framework. The algorithm of a series of actions such as dataset preparation, image preprocessing and augmentation, model architecture and training and performance evaluation.

A. Dataset Preparation

A heterogeneous collection and curation of microscopic peripheral blood smear images were collected and curated to contain leukemic and normal cell samples with a wide variation of morphological variations. The images were thoroughly annotated by professional experts with the leukemic cells identified by means of bounding boxes to make sure that the model is trained correctly and that no bias was introduced during testing.The data was divided into three groups to provide consistent model training and balanced evaluation. The training set was of about 88 per cent (6600 images), validation and test sets were 8 and 4 per cent respectively (623 and 312 images). On the training set, the underlying features of leukemia cells were learned and the training progress was optimized on the validation set. The test set was not observed during training but was simply used to test the model lastly and give objective measurement of the performance of the model.

B. Data Augmentation and Image Preprocessing.

To improve the quality of the input data and to improve the model, many preprocessing methods were applied. To begin with, all images were resized to the same size that they can be utilized in the YOLOv8 network as its input. Image normalization was performed to put the pixel values within a constant range that aids in stabilizing the training process. Moreover, morphological features and contrast enhancement methods were used to enhance the visibility of cells and provide more information on the morphological features. Different data augmentation strategies were used in order to make the datasets more diverse and enhance the ability to generalize. These were image rotation, horizontal and vertical flipping, scaling, contrast adjustment and augmentation of mosaic. These changes enable the model to acquire the characteristics of invariance and are able to identify the leukemia cells in different staining, illumination, and cell types.

C. YOLOv8 Model Architecture and Training.

YOLOv8 object detection framework has been picked because it has great accuracy in detection and it can provide real time inference. YOLOv8 is a single-stage model in which both object classification and location are carried out at the same time, which makes it effective in clinical screening.

YOLOv8 architecture has three main components that include backbone, neck, and detection head. The backbone network does the extraction of hierarchical features in the input images. It employs the convolutional layers to extract both low-level features of edges, textures, and high-level semantic features of the leukemic cell structures. The neck block combines both a Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to provide better multi-scale feature fusion. The design also enables the model to identify objects of various sizes which is mostly critical in the identification of leukemia cells which vary in shape and size.

The final prediction tasks are carried out in the detection head, namely, the regression of bounding box, object classification, and the confidence score. It determines the position of the leukemic cells in the image and gives it probability scores that represent the chances of the object detected in the image being of the leukemia type. Transfer learning was also used to speed up convergence of the model and enhance the accuracy of detection by initializing the model with pretrained weights. The network was subsequently trained on annotated images with hyperparameters that have been optimized like the learning rate, batch size and the number of epochs.

The grooming task maximized a compound loss that consists of:Bounding Box Regression Loss to learn an object localization correctly. Classification Loss In order to make the right selection of the classes. Focal Loss Distribution Focal Loss (DFL) to improve the precision of bounding box regression. All these contribute to the proper cell detection and classification of leukemia..

D. Performance Evaluation

The evaluation of the model after training was done with the unseen test dataset. The measure of the proposed system performance was based on standard object detection metrics. Such metrics are: Precision ,Recall , F1-score. Mean Average Precision (mAP). Besides the measures of accuracy, the inference time was also plumbed to determine the procedure ratio of the model. This assists in establishing the ability of the proposed system to be implemented in the real-time clinical screening settings.

E. System Architecture

Fig:1 proposed leukemia detection system has the overall architecture that is structured to follow an automated pipeline of leukemic cells identification in real-time. To start with, to enhance the quality of the input, microscopic blood smear images are first obtained and undergo a preprocessing and augmentation process to diversify the datasets. The improved images are subsequently inputted to the YOLOv8 network with the backbone identifying deep spatial characteristics, the neck carrying out multi-scale characteristics merger, and the sensing head forecasting boundings boxe.

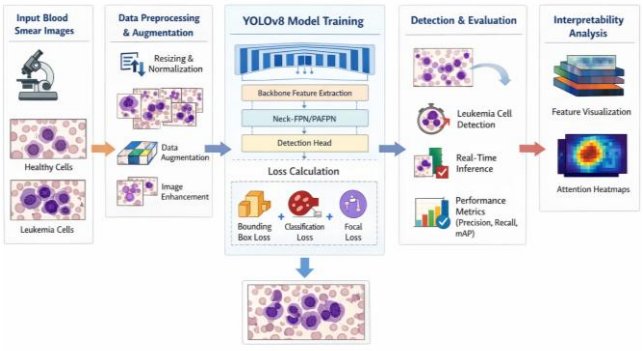

Fig:1 system architecture

The drilled model are used during the process of inference, which examines every picture that has not been seen and generates a bounding box around the potential leukemia cells and a confidence score. Detection performance is then assessed by the system by means of conventional metrics and visual interpretability methods. This end-to-end system allows rapid, precise, and validated detection of leukemia cells and also lessens the workload of manual diagnostic machines as well as helping hematologists in mass screening.

IV. Result and Discussion

The training and evaluation data set is comprised of microscopic peripheral bodily fluid into four classes as in fig:2. The distribution of the labeled cases indicates that the Pre stage has the most number of samples of about 38,000 cases, then the Early stage cases are about 14,500 cases, Benign cases are about 10,500 cases, and Pro cases are about 9,700 cases. This distribution suggests a moderate imbalance in classes in which Pre class prevails in the data.

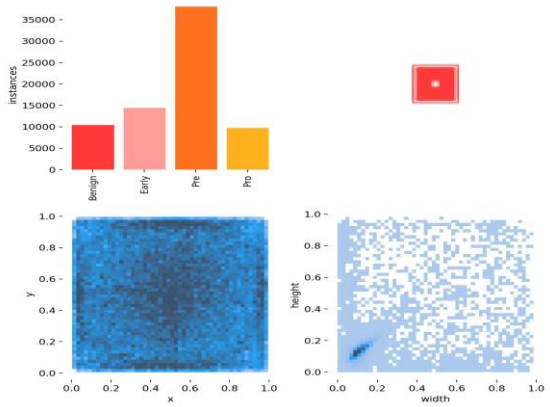

Fig:2 Dataset Distribution Analysis

Nonetheless, training rectified this imbalance through data augmentation methods and balanced sampling plans to make sure that the model acquires the representative features of every category. It can be seen that the bounding box distribution analysis indicates that the bulk of cell annotations is concentrated in the middle part of the images, which is the primary field of view in microscopic blood smear imaging.

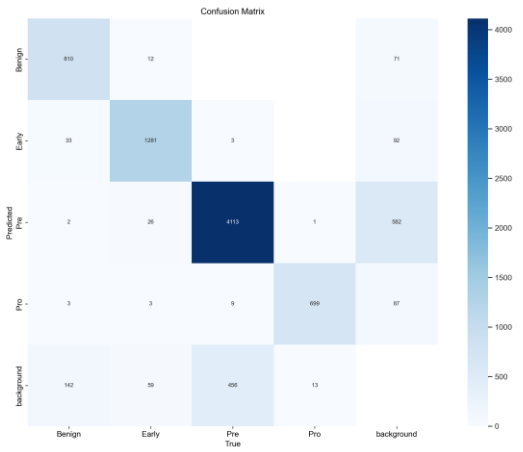

Fig:3 Confusion Matrix Analysis

The array of disorder as in fig:3 offer an elaborated insight into the performance of the trained model in the four leukemia classes with respect to classification. The results indicate that the exemplary aright categorised 810 samples of the Benign class, 1281 samples of the Early class, 4113 samples of the Pre class, and 699 samples of the Pro class. The highest number of correct predictions was observed for the Pre class (4113), which can be attributed to its larger number of training samples and distinctive morphological characteristics. Despite the overall strong performance, a small number of misclassifications were observed. For example, 12 Benign samples were misclassified as Early, and 71 Benign samples were predicted as background. Similarly, 33 Early samples were misclassified as Benign, 3 as Pre, and 92 as background. In the case of the Pre class, 26 samples were incorrectly predicted as Early, and 582 were categorized as background.

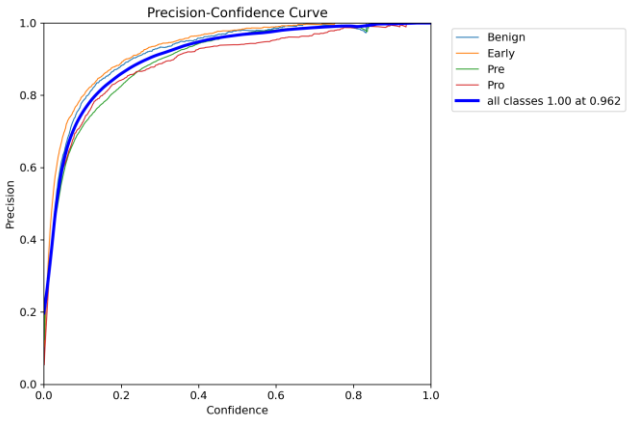

Fig:4 Precision–Confidence Analysis

Precision-confidence curve in fig:4 depicts how change in prediction precision at varying levels of confidence occurs. These findings demonstrate that the degree of accuracy gradually rises with the degree of the confidence threshold meaning that the model is more selective in its predictions. The model had a maximum accuracy of 1.00 at the confidence level of about 0.962 that is, almost all the detections at the confidence level above this are true positive predictions. Such behavior is of specific interest in clinical practice when one needs to ensure as low false positives as possible in order to be sure of a diagnosis. The curve also suggests that the Early and Pro classes have higher accuracy with a large range of confidence thresholds which suggests that the model has acquired good discriminative features to these stages of leukemia.

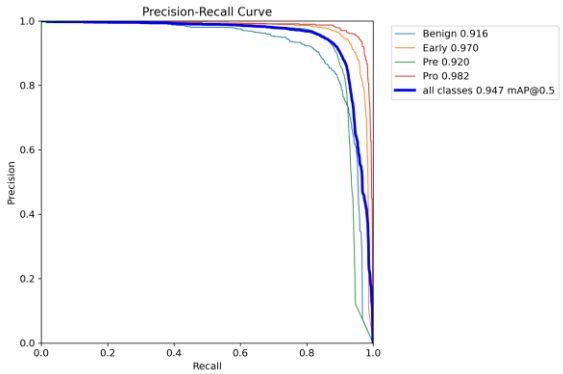

Fig:5 Precision–Recall Performance

preprocessing, augmentation and transfer learning to increase detection accuracy/robustness. The performance of the experiment is shown to be high, with a precision of about 0.97, a recall of 0.97, and mAP at 0.5 of 0.947, meaning that the model can localize and classify leukemia cells between various stages with high precision and recall. Quantitative measurements of the method as well as visual images of the method of detection affirm the accuracy of the method even in the complex microscopic images that may have thick or partner cells. In general, the proposed approach will offer a rapid, accurate, and practical solution that can assist hematologists in the screening of early-stage leukemia, decrease the number of manual operations and add to the efficient diagnostic processes in a clinical setting..

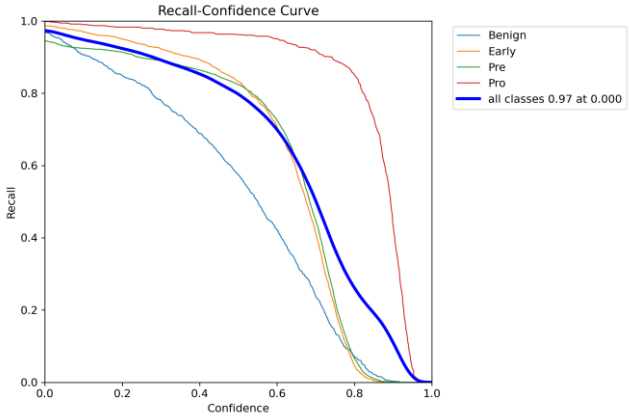

Fig:6 Precision–Confidence Analysis

The recall–confidence curve evaluates as in fig:6 how well the model detects leukemia cells across numerous confidence levels. The findings indicate that the overall recall is approximately 0.97 at the lowest confidence threshold as this means that the model has been able to identify approximately 97 percent of the leukemia cells in the dataset. Given a higher confidence threshold, recall will decline gradually since the more detections will not pass the larger confidence threshold. In contrast, the Benign class exhibits slightly lower recall, likely due to morphological similarities between benign cells and early-stage leukemic cells, which can occasionally lead to missed detections.

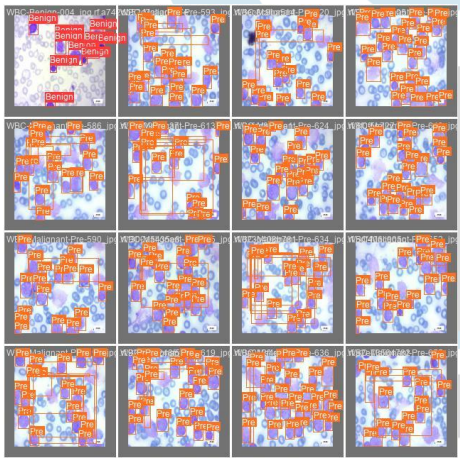

Fig:7 Detection Visualization Results

The qualitative results of detection also prove the efficiency of the suggested YOLOv8-driven detection system as in fig:7. The visualization pictures indicate that the model is able to identify several leukemic cells in only one microscopic image and correctly localize a binding box around the cells found in the image. The model also properly classifies the identified cells based on their respective classifications such as Pre, Early, Pro and Benign. These detection results suggest that the model has a high level of performance even in the complicated environment where cells are closely spaced, a bit of overlap or are of different sizes and orientations. The model is also resistant to changes in staining conditions and illumination levels that are frequent in microscopic blood smear images. These qualitative findings affirm the practicability of the system to automated leukemia detection.

TABLE1. COMPARATIVE PERFORMANCE

Model	Precision	Recall	mAP / Accuracy
CNN (Traditional)	0.88	0.86	0.87
Faster R-CNN	0.92	0.90	0.91
SSD (Single Shot Detector)	0.90	0.89	0.90
YOLOv5	0.94	0.93	0.93
Proposed YOLOv8	0.97	0.97	0.947 (mAP@0.5)

The comparative analysis as As can be seen in table:1, the suggested YOLOv8-based leukemia detection model is more successful than various other deep learning models currently available. The conventional CNN-based approaches have decent classification accuracy, but cannot localize the leukemia cells in the microscopic images. Two-stage detectors like Faster R-CNN have high detection accuracy, but have lower inference rates, so they can not be used in real-time clinical screening. Single-stage detectors such as SSD and YOLOv5 are faster in terms of detection, but they are a little bit less accurate in detecting small leukemia cells. The presented YOLOv8 model has increased precision (0.97), recall (0.97), and mAP@0.5 (0.947) and proves to be more superior in detection but can still perform inference in real-time. This contributes to the fact that this system is more appropriate. automated leukemia screening and clinical decision support.

V. CONCLUSION

In study, a YOLOv8-based profound acquisition framework was formulated for the automated sensing of leucaemia cells from research peripheral blood smear images. The projected system effectively combines data preprocessing, augmentation and transfer learning to increase detection accuracy/robustness. The performance of the experiment is shown to be high, with a precision of about 0.97, a recall of 0.97, and mAP at 0.5 of 0.947, meaning that the model can localize and classify leukemia cells between various stages with high precision and recall. Quantitative measurements of the method as well as visual images of the method of detection affirm the accuracy of the method even in the complex microscopic images that may have thick or partner cells.Overall, the suggested solution will provide a quick, precise, and feasible solution that can help hematologists in the screening of the early-stage leukemia, reduce the number of manual operations and contribute to the efficient diagnostic procedures in a clinical environment.

REFERENCES

[1] C. S. Vomanesh, S. R. Kumar, and P. R. Reddy, "Artificial intelligence–based leukemia diagnosis using complete blood count data," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 321–329, 2023.

[2] Q. Fan, Y. Zhang, H. Li, and X. Wang, "QCResNet: A deep convolutional neural network for acute lymphoblastic leukemia classification from peripheral blood smear images," *Computers in Biology and Medicine*, vol. 156, pp. 106–115, 2023.

[3] A. Y. Khan, M. A. Rahman, and S. Ahmed, "Automated leukemia detection using improved convolutional neural networks for microscopic blood cell classification," *Biomedical Signal Processing and Control*, vol. 80, pp. 104–112, 2023.

[4] A. Ramagiri, P. V. Kumar, and R. S. Reddy, "Comparative analysis of machine learning and deep learning techniques for leukemia detection using blood smear images," *Journal of Medical Systems*, vol. 47, no. 5, pp. 1–12, 2023.

[5] M. Ashok, K. Srinivasan, and R. Babu, "Automated detection of acute lymphoblastic leukemia using computer vision and machine learning techniques," *IEEE Access*, vol. 11, pp. 43124–43135, 2023.

[6] Zarish, M. Iqbal, and A. Rehman, "Machine learning approaches for predicting survival rates in leukemia patients using SEER clinical data," *Healthcare Analytics*, vol. 2, pp. 100091, 2022.

[7] J. Sheet, S. Chatterjee, and S. Roy, "Transfer learning-based approach for leukemia detection using MobileNetV2 and convolutional neural networks," *Expert Systems with Applications*, vol. 213, pp. 119–128, 2023.

[8] K. K. Jha, R. Kumar, and P. Singh, "Ensemble deep learning model for leukemia detection using augmented medical image datasets," *Artificial Intelligence in Medicine*, vol. 124, pp. 102–111, 2022.

[9] R. Arif, M. U. Akram, and A. Khan, "A deep learning framework for acute lymphoblastic leukemia detection using preprocessing and CNN-based classification," *Journal of Biomedical Informatics*, vol. 126, pp. 103–112, 2022.

[10] A. M. Basymeleh, M. A. Alsharif, and A. Alqahtani, "Multi-class leukemia classification using transfer learning with convolutional neural networks," *Diagnostics*, vol. 12, no. 9, pp. 1–15, 2022.

[11] T.-Y. Lin *et al.*, "Feature Pyramid Networks for Object Detection," *Proc. IEEE CVPR*, pp. 2117–2125, 2017.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[13] R. Girshick, "Fast R-CNN," *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.

[14] L. Deng and D. Yu, "Deep Learning: Methods and Applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[15] World Health Organization, "Leukemia: Early Diagnosis and Treatment Guidelines," WHO Press,