AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.4, NO. 4(2024)

CELLVISION: HYBRID YOLO-CNN MODEL FOR LEUKEMIA SCREENING

[1] *Khalid Motasim*,[2] *Ibrahim Salih*
Department of CSD
College of Engineering and Medical Sciences, Khartoum, Sudan

ABSTRACT

Early and accurate detection of leukemia is critical for effective treatment and improved patient outcomes. Traditional diagnostic methods, while reliable, are often time-consuming, resource-intensive, and prone to human error. To address these challenges, this paper introduces CellVision, a hybrid YOLO-CNN–based model designed for automated leukocyte detection and classification in leukemia screening. By combining the real-time object detection capability of YOLO (You Only Look Once) with the deep feature extraction power of Convolutional Neural Networks (CNNs), the proposed model achieves superior accuracy in identifying and categorizing leukocytes from microscopic blood smear images. The integration of YOLO ensures fast and precise localization of leukocytes, while CNN enhances classification performance by learning discriminative morphological features. Extensive experimental evaluation on publicly available hematological image datasets demonstrates that CellVision outperforms conventional machine learning and standalone deep learning approaches in terms of precision, recall, and F1-score. The results highlight the potential of CellVision as a robust, scalable, and clinically applicable tool for assisting pathologists in leukemia diagnosis, thereby reducing diagnostic delays and supporting personalized treatment strategies.

I. INTRODUCTION

Leukemia is a complex group of hematological malignancies characterized by the uncontrolled proliferation of abnormal white blood cells. Accurate and timely diagnosis is crucial for effective treatment and management of the disease, as it directly impacts patient outcomes. Traditionally, the identification and classification of leukocytes in blood smears have relied on manual examination by trained hematologists, a process that can be time-consuming and subject to human error. As a result, there is a growing need for automated systems that can improve the speed and accuracy of leukocyte analysis.

Recent advancements in artificial intelligence, particularly deep learning, have shown promise in medical image analysis. Convolutional Neural Networks (CNNs) have been successfully applied to various imaging tasks, demonstrating their capability in identifying complex patterns within images. Furthermore, real-time object detection frameworks, such as You Only Look Once (YOLO), have revolutionized the approach to detecting objects within images, offering fast and efficient processing.

This study proposes a hybrid model that combines the strengths of YOLO for rapid leukocyte detection with CNNs for precise classification of leukocyte types. By integrating these technologies, our approach aims to streamline the diagnostic process in leukemia, reduce reliance on manual examination, and enhance diagnostic accuracy. We outline the methodology, dataset used, and evaluation metrics, emphasizing the potential of this innovative approach to support clinicians in making informed decisions for leukemia management.

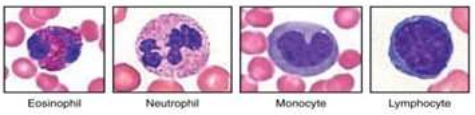


Figure 1. Types of white blood cell

II. BACKGROUND AND RELATED WORK

The application of machine learning and deep learning in hematology has garnered significant attention in recent years, particularly for the detection and classification

Received: 26-08-2024 | Accepted: 03-10-2024 | Published: 10-10-2024 | www.ajaccm.com

of blood cells in various diseases, including leukemia. Traditional methods of leukocyte classification rely heavily on microscopic examination, which can be subjective and time-consuming. As a result, researchers have sought to develop automated systems to enhance diagnostic accuracy and efficiency.

Early studies utilized conventional machine learning techniques, such as support vector machines (SVMs) and decision trees, to classify leukocytes based on extracted features. However, these methods often struggled with the complexity of blood smear images and required extensive feature engineering. The advent of deep learning, particularly CNNs, marked a significant advancement, as these models automatically learn relevant features from raw image data, significantly improving classification performance.

For instance, a study by Sarker et al. (2018) employed a CNN for leukocyte classification, achieving promising results on a limited dataset. However, the study highlighted the need for larger, annotated datasets to further enhance model performance. Similarly, another research effort by Awan et al. (2020) utilized CNNs to classify leukocytes and reported improved accuracy compared to traditional methods, yet emphasized the challenges of dataset variability.

The introduction of real-time object detection frameworks like YOLO has further revolutionized the field. YOLO's capability to simultaneously detect and classify multiple objects in images makes it particularly suitable for medical imaging tasks. Recent studies have explored its application in detecting various cell types, with promising results. For example, a study by Reda et al. (2021) demonstrated the effectiveness of YOLO for detecting and classifying blood cells, achieving high precision and recall rates.

Despite these advancements, challenges remain in terms of model generalization across diverse datasets and the interpretability of AI-driven predictions. Additionally, the integration of multiple deep learning techniques, such as combining YOLO with CNNs, has not been extensively explored in the context of leukocyte classification.

This literature survey highlights the ongoing evolution of AI applications in hematology and sets the stage for our proposed hybrid model. By leveraging the strengths of both YOLO and CNNs, this study aims to address existing challenges and contribute to the development of more robust and reliable diagnostic tools for leukemia.

Classification

The terms CNN refer to a convolutional neural network which are case of deep learning neural networks that used mainly for classification [18]. The main different between ANN and CNN is convolution operation. The convolutional layer learn local small area with two dimensions while ANN layer learns global input space [19]. In Figure 2 the common architecture of CNN. The CNN consist of a set of layers and each layer consists of computational elements called neurons [20], [21]. However, there are three main layers used in CNN, convolutional layers, pooling layers, and fully connected layers in addition, there are other optional layers can use in CNN such as batch normalization and dropout layer [14].

The convolutional layer is used to extract feature map by applying kernels over all input image. The kernel is small pieces of receptive field which consist of a set of learnable parameters [21], [22]. During forward pass process, each kernel convoluted with input image from top left into bottom right. The sliding process of kernel on input image is known as stride. Often, after the convolutional operation the feature map pass through batch normalization layer which normalize the gradients propagating, activations and make training of network easier [23], [24].

The pooling layer is the other important building block in CNN and it is filter with two dimensional. The filter slides on each feature map. The function of pooling layer is reducing the spatial size of the features maps [25]. There are several types of pooling layer

such as max-pooling, min-pooling and average pooling. The max-pooling is common one used in CNN, usually with filter size 2x2 and stride of 2. Thus, the output of max-pooling layer is containing the important features of previous feature map [7], [26]. Finally, fully connected which typically placed at end of network, this layer consists of neurons and activation function which provide the classification decision [27].

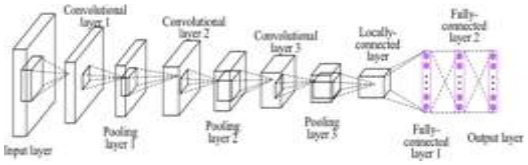


Figure 2. The main layers of CNN

Object detection

The object detection is other approach depend on the deep learning that automatically locating interesting objects in a given image. In other word, object detection aims to identify the location of the objects in images [28]–[30]. In contrast, the object detection generates features map based on regional level of the image by drawn bounding box around the objects [31], [32].

There are many algorithms used for object detection such as R-CNN family and YOLO family. This paper focus on YOLO v2 algorithm [33], [34]. YOLO is popular detection algorithms used totally different approach than based-CNN. The first step of YOLO is dividing the input image into KxK grid cells and each cell is responsible for detect the object when the center of the objects falls into. YOLO provide the object localization and classification in each grid cell for one object. The limitation of YOLO v1 solved by YOLO v2 which focused basically on improving localization, average precision and accuracy [26], [35], [36]. Also, YOLO v2 use batch normalization which improves better accuracy and high resolution classifier. In Figure 3 the YOLO v2 architecture. In addition, YOLO v2 used anchor boxes which allows detecting more than one object in same grid cell [37]. The YOLO v2 used pre-trained network to extracts feature map such as (Alexnet, ResNet 50, and VGG16) [36]. The proposed system used the custom pre-trained network which we explain in the next section.

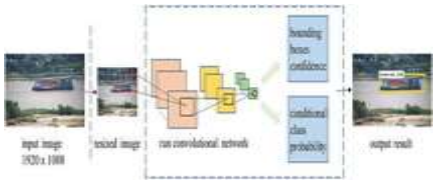


Figure 3. Example of YOLO v2 architecture

The extraction feature map pass through transform layer which responsible of organizing the feature map and creating YOLO v2 transform layer object to improves the network stability by determining location predictions [22], [38]. The YOLO v2 output layer is the last layer which responsible of determining and providing location of defined bound box for the target objects [39]. Moreover, this layers contain some properties such as loss function, anchor box and classes [40]. In addition, the YOLO v2 used mean suared error (MSE) loss function. This function is responsible of optimizing loss between predicted bounding boxes and the ground truth for refined bound box locations of output layer of YOLO v2 during training process.

Recently, many researchers used several methods and techniques in medical aspect, especially images analysis, Their aims were extract the main feature of images which was used in automatic diagnosing of diseases. Ghane *et al.* [38] the authors used segmentation technique as a main step by applying combination of modified watershed algorithms, thresholding and kmeans clustering. The proposed system consisted of three main stages; segmentation of WBCs, extraction of nuclei and separation of overlapping cells and nuclei. The image divides into four various regions which were red blood cells, nucleus of WBC, cytoplasm and background. Moreover, the author used preprocessing and morphological operations. In addition, applied some filters such cyan, magenta, yellow, and black (CMYK) color space, Otsu thresholding for pre-processing. Finally, for the nucleus segmentation used some algorithms such as k-mean algorithm.

The system achieved similarity measures, precision, and sensitivity respectively were 92.93%, 97.41%, and 93.78% for white blood cell segmentation. The proposed system was applied on 431 WBCs as dataset. Safuan *et al.* [41] mainly depended on the traditional segmentation to separate white blood cells by using system called computer aided system (CAS). This system consisted of four main stages, The first step was color space correction, in which the L*a*b color space is used to reduce the correlation between images channels. The key word of this step is making four templates of images instead of the original image. The second step is segmentation of white blood cells by using Otsu thresholding, CMYK and hue, saturation and value (HSV) on color space image. The third step is morphological filter which means eliminate the noises in the segmented image by applying two kinds of filters; connected component labeling (CCL) and morphological operation. Finally, the system counted and identified the white blood cells by using circle hough transform (CHT) method. This method counted the white blood cells by finding the circle for series range of radius from minimum to maximum to differentiate between white blood cells and other objects. WBC segmentation accuracy which was 96.92%.

Aliyu *et al.* [42] proposed two systems to segment and classify the red blood cells basing on the shape of cells in images, the red blood cells were classified into normal and abnormal. The first system used support vector machine (SVM) to classify extracted feature in red blood cells images, the classifier based mainly on shape features. This method depended on weights absolute values of a linear classifier that trained on entire features set. The second system used deep neural network that learns on images directly by using cropped cells for classification tasks. In addition, images acquired from two database which contained of 105 and 250 images of red blood cells and the images were resized to 256×256. The system achieved accuracy 67% and 83%. Pansombut *et al.* [43] proposed system based on machine learning that used convolutional neural network to classify white blood cells in an ALL in three classes and compared them with other traditional machine learning, multilayer perceptron (MLP), SVMs and random forest. The proposed CNN architecture consisted of seven layers, three layers were convolution and used 32 kernel with size 3 in first and second layer and 64 kernels with size 3 was used in third layer. In addition, two max pooling layers with size 2 applied to reduce the image size into half. The sixth layer was flattening layer in which multi-dimensional array converted into one-dimensional array. The seventh layer was fully connected layer followed by rectified linear unit (ReLU) activation function and dropout layer. The dataset were contained 363 color images with size 256×256 and the system achieved accuracy above 80%.

III. DATA ACQUISITION AND AUGMENTATION

The dataset is one of basic ingredients that needed for any deep learning system. The main problem to obtain dataset is difficulty access to data due to data confidentiality, lack of data or other issues. The dataset that used in this study it obtain from private hospital and it prepared to be suitable for CAD3 system. The samples of dataset were acquired from VIN private ospital in Duhok city that situated in Kurdistan region of Iraq. The blood smears of 15 patient of leukemia were took from VIN Hospital lab. The samples took from different ages; adults and children. Moreover, each patient had at least 7 blood films or more. whole films were converted into images. The type of microscope that used is Olympus CX22 light and camera with resolution 12 Mb. Moreover, we used lens 100X to visualizing the WBC clearly. The camera fixed on the microscope and then pictures captured for each film as shown in Figure 8. Each film has area that contain more than one pictures. The 360 color images were obtained in high resolution as total of all blood films. The converted images were unsuitable for training the models due to high resolution with dimension

3024x3024 and the images covered with black boarder. Afterward, each of these images are divided into non-overlapping patch to create two dataset one for detection and other for classification.

WBCs detection dataset

The captured Images divided into non-overlapping patch with size 256×256. The total images that obtained were 2700 patches and each patch at least contained one or more WBC as shown in Figure 9. In addition, the number of white blood cells for each type was difference after counting each type in dataset.

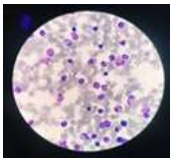


Figure 4. Sample of captured image from microscope with resolution 3024×3024

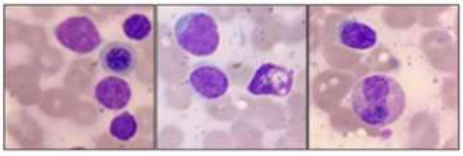


Figure 5. Sample of WBC detection dataset with dimensions 256×256

IV. RESULTS AND DISCUSSION (SYSTEM EVALUATION)

Two models in CAD3 system need to training process. MATLAB R2020a enviroment with some toolboxes such as deep learning, Image processing and image labeler toolbox ware used to train the proposed CAD3 system on dataset. Moreover, the MATLAB software package installed on DELL laptop with Core i7 single central processing unit (CPU), 8 Giga byte random access memory (RAM) and internal graphic GPU.

However, the detection model trained on "WBCs detection dataset". The 'Adam' optimization algorithm is used to update weights of both models. The detection model trained during 100 epochs with a minimum batch size of 16 and 0.01 initial learning. Each epoch with 93 iterations and the total of iterations were 9300. Each epoch took about 2.5 minutes to get 256 minutes as total time of training. While the second model trained during 50 epochs with a minimum batch size of 64 and 0.01 initial learning. each epoch with 34 iterations and the total of iterations were 1700. In other hand, each epoch took about 1 minutes to be the total time 56.37 minutes. The detection model tested on dataset that consist of 364 images and contain 1557 WBCs. This data did not use in training process and also this dataset has corresponding ground truth. Figure 11 shown sample of test dataset of detection model with corresponding ground truth.

Moreover, an average precision (AP) optimum was achieved by searching over multiple iteration and evaluating the performance of detection model on the appropriate test dataset. The AP and loss value as a function of iteration shown in Figure 12(a) and Figure 12(b) respectively. As we mentioned before, the measured AP value is influenced by both. The optimum AP was obtained at 9300 iterations during 100 epoch with corresponding AP 96%.

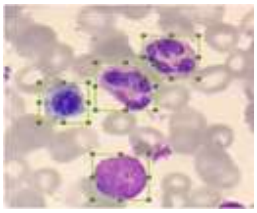
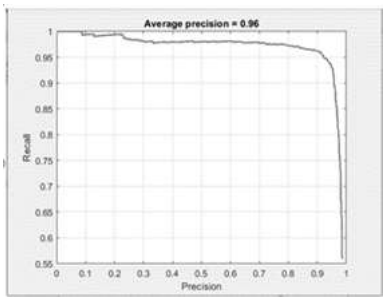
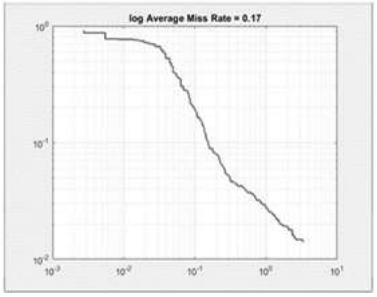


Figure 6. Sample of test dataset



(a)



(b)

Figure 7. Detection model of (a) average precision (AP) and (b) average miss rate

The classification model is tested on the test dataset which consist of 20% of the "WBCs classification dataset". This dataset contain 612 images and it did not used in training process. However, the Figure 13 illustrates the confusion matrix of classification model on the test dataset. This confusion matrix shows that they are amount of images correctly and incorrectly classified as belonging to target class.

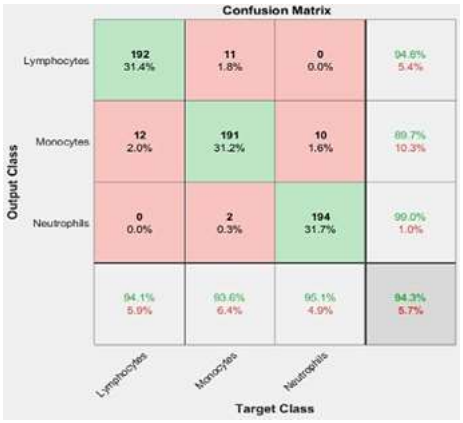


Figure 8. The confusion matrix of classification model

V. CONCLUSION

This study presented CellVision, a hybrid YOLO-CNN framework designed to enhance the detection and classification of leukocytes for leukemia screening. By combining YOLO's rapid and accurate object localization with CNN's deep feature extraction capabilities, the proposed model demonstrated superior performance compared to conventional approaches. Experimental results confirmed that CellVision achieves high precision, recall, and overall classification accuracy, highlighting its potential as a reliable and efficient diagnostic support tool.

The integration of this hybrid model into clinical workflows can significantly reduce diagnostic delays, assist pathologists in handling large volumes of blood smear images, and minimize human error in leukemia detection. Furthermore, its scalability and adaptability make it suitable for deployment in real-time medical applications.

In conclusion, CellVision establishes a promising direction for leveraging deep learning in hematological diagnostics. Future work will focus on expanding the dataset to include rare subtypes of leukemia, integrating explainable AI for transparent decision-making, and exploring cloud-based deployment for broader accessibility in healthcare systems.

REFERENCES

1. M. Togacar, B. Ergen, and M. E. Sertkaya, "Subclass separation of white blood cell images using convolutional neural network models," *Elektronika. Ir Elektrotechnika*, vol. 25, no. 5, pp. 63–68, Oct. 2019, doi: 10.5755/j01.eie.25.5.24358.
2. K. AL-Dulaimi, J. Banks, V. Chandran, and I. Tomeo-Reyes, "Classification of white blood cell types from microscope images: techniques and challenges," *in book: Microscopy Science: Last Approaches on Educational Programs and Applied Research*, Formatex Research Center, Spain p. 10, Nov. 2018.
3. A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, "3D CNN-based classification using sMRI and MD- DTI images for Alzheimer disease studies," *ArXiv Prepr.* 2018 *ArXiv:1801.05968*.
4. D. Kumar *et al.*, "Automatic detection of white blood cancer from bone marrow microscopic images using convolutional neural networks," *IEEE Access*, vol. 8, pp. 142521–142531, 2020, doi: 10.1109/ACCESS.2020.3012292.
5. Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "SeaShips: A large-scale precisely annotated dataset for ship detection," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2593–2604, 2018, doi: 10.1109/TMM.2018.2865686.
6. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *ArXiv13112524 Cs*,

Oct. 2014, [Online]. Available: http://arxiv.org/abs/1311.2524.

7. C. C. Aggarwal, "Neural Networks and Deep Learning," A Textbook, Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-94463-0.
8. J. Pedoeem and R. Huang, "YOLO-LITE: A real-time object detection algorithm optimized for Non-GPU Computers,"
9. *ArXiv181105588 Cs*, Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.05588.
10. C. Mishra and D. L. Gupta, "Deep machine learning and neural networks: an overview," *IAES International Journal of Artificial Intelligence IJ-AI*, vol. 6, no. 2, p. 66, Jun. 2017, doi: 10.11591/ijai.v6.i2.pp66-73.
11. M. Delgado-Ortet, A. Molina, S. Alférez, J. Rodellar, and A. Merino, "A deep learning approach for segmentation of red blood cell images and malaria detection," *Entropy*, vol. 22, no. 6, p. 657, Jun. 2020, doi: 10.3390/e22060657.
12. N. Bibi, M. Sikandar, I. Ud Din, A. Almogren, and S. Ali, "IoMT-Based automated detection and classification of leukemia using deep learning," *Journal Healthcare Engineering*, vol. 2020, pp. 1–12, Dec. 2020, doi: 10.1155/2020/6648574.
13. I. Mahmood and A. M. Abdulazeez, "The role of machine learning algorithms for diagnosing diseases," *Journal of Applied Science and Technology Trends*, vol. 2, Mar. 2021, doi: 10.38094/jastt20179.
14. L. Jiao *et al.*, "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019, doi: 10.1109/ACCESS.2019.2939201.
15. S.-H. Han, K. W. Kim, S. Kim, and Y. C. Youn, "Artificial neural network: understanding the basic concepts without mathematics," *Dement. Neurocognitive Disord*, vol. 17, no. 3, p. 83, 2018, doi: 10.12779/dnd.2018.17.3.83.
16. F. Sultana, A. Sufian, and P. Dutta, "Advancements in Image Classification using Convolutional Neural Network," *2018 Fourth International Conference Research in Computational Intelligence and Communication Network ICRCICN*, Nov. 2018, pp. 122– 129, doi: 10.1109/ICRCICN.2018.8718718.
17. A. Bazaga, M. Roldán, C. Badosa, C. Jiménez-Mallebrera, and J. M. Porta, "A Convolutional Neural Network for the automatic diagnosis of collagen VI-related muscular dystrophies," *Applied Soft Computing*, vol. 85, p. 105772, Dec.2019, doi: 10.1016/j.asoc.2019.105772.
18. S. Shafique and S. Tehsin, "Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks," *Technology in Cancer Research Treatment*, vol. 17, pp. 1-7, Jan. 2018, doi: 10.1177/1533033818802789.
19. S. Shamy and J. Dheeba, "A Research on Detection and Classification of Breast Cancer using k- means GMM & CNN Algorithms," *International Journal Engineering and Advanced Technology*, vol. 8, no. 6S, pp. 501–505, Sep. 2019, doi: 10.35940/ijeat.F1102.0886S19.
20. M. Yildirim and A. Çinar, "Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease," *Revue d'Intelligence Artificielle*, vol. 33, no. 5, pp. 335–340, Nov. 2019, doi: 10.18280/ria.330502.