Received: August 4, 2024. Revised: November 2, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

MCNet: Mask Cell of Multi Class Deep Network for Blood Cells Detection and Classification

Ibtihaal M. Hameed[1] **Mohammed Al-Mukhtar**[2] **Ammar S. Al-Zubaidi**[2]*

[1]*University of Baghdad, Ibn Sina Center for E-Learning, Baghdad, Iraq*
[2]*Computer Centre, University of Baghdad, Baghdad, Iraq*
* Corresponding author's Email: Ammar.Sabah@cc.uobaghdad.edu.iq

Abstract: Physicians are likely to expend significant labor and time while manually calculating blood smears. Automatic computer-based methods for classifying acute lymphoblastic leukemia have trouble correctly lighting stained white blood cell microscopy images and accurately separating cells that touch or overlap. Additionally, incorporating machine learning techniques into medical services is very hard because doctors can deal with rough guesses as long as the results aren't too bad, but they can't use these calculations for actual medical care. Enabling a deep network to have knowledge of the accuracy of its own predictions is a fascinating and crucial issue. Most instance segmentation frameworks weigh the mask quality during the instance segmentation process based on classification confidence. Here, we consider the context of this problem and present Mask Cell of multi class deep network (MCNet) as a new network that has the module to learn about the quality of the predicted instance masks. Our proposal entails using faster R-CNN, such as segmentation on white blood cell microscope images, to accurately categorize acute lymphoblastic leukemia cases. This approach aims to enhance the efficiency and effectiveness of the diagnostic process. The suggested network block combines the instance feature with the matching anticipated mask to estimate the proposed mask IoU. In this work, we used the transfer learning approach to apply Mask R-CNN to segment white blood cells on a microscope image. To address the issue of poor lighting in stained white blood cell microscopy pictures, we included a contrast enhancement procedure in the image dataset. The comparative experiment applies YOLO v9 for classification and Mask R-CNN. The MCNet approach adjusts the discrepancy between the quality of the mask and its proposed detection, enhancing the effectiveness of instance segmentation. The final results for two datasets trained using PBC and BCCD are as follows: the accuracy of mAP@IoU0.50 for the PBC dataset is 95.70, while the accuracy for the BCCD dataset is 96.76, with recall and precision both coming in at 97.23 and 96.72 respectively.

Keywords: YOLO 9, R-CNN, Blood smear detection.

1. Introduction

Blood supplies nutrition and oxygen to the live cells found in many organs and tissues. It removes the waste so that detoxification can occur. It controls body temperature and delivers hormones to their intended locations of action to battle infections [1]. A peripheral blood smear is a standard laboratory test that offers the doctor comprehensive information on a patient's overall condition. It provides a statistical and qualitative assessment of blood components, including primary cells and platelets.

The colored liquid of blood which is called Plasma forms 55% of the whole blood while the cells make up 45% [2]. Human blood comprises three types of blood cells: white blood cells (leukocytes), red blood cells (erythrocytes), and platelets (thrombocytes) [3]. Erythrocytes are the most abundant blood cells and contain the hemoglobin protein, which gives the cell its red hue and facilitates oxygen transport from the lungs to the body's tissues. Thrombocytes, also referred to as platelets, are smaller in size compared to erythrocytes and do not possess a nucleus. They play a crucial role in the formation of blood clots and the prevention of bleeding [4].

White blood cells (WBC) are non-granular, without color, and consist of a varying number of nuclei surrounded by a small quantity of cytoplasm [4]. The lymphatic system contains a large number of leukocytes. WBCs play a crucial role in the immune system by protecting the body from infectious illnesses and external threats in the bloodstream. Lymphocytes, eosinophils, neutrophils, basophils, and monocytes are the five distinct kinds of nucleated cells comprising leukocyte [5]. Pathologists will be able to tell the difference between several blood diseases and illnesses, including anemia, leukemia, and malaria, if they can correctly identify, count, and group WBC and see how the percentages change between subtypes. Enhanced information will facilitate therapies, mitigate detrimental medication interactions, and oversee the patient's health [6].

Certain diseases can cause changes in the appearance or number of blood cells, so blood cell detection is crucial for the diagnosis and treatment of blood and diseases [2]. Blood cell detection has become an important auxiliary tool in the medical field [7]. Pathologists manually identify blood cells, a process that is sometimes susceptible to human error and produces incorrect findings. Pathologists may use this method with variations both within and between various courses, resulting in a laborious and time-consuming process [8]. In the conventional approach for identifying WBC, RBC, and platelets, distinguishing them from other components of blood is challenging because of their comparable textures and irregular borders. Furthermore, WBCs exhibit a diverse array of color, shape, form, and intensity [9]. Varied staining and illumination circumstances further complicate the identification of white blood cells. Furthermore, the inclusion of multi-class categorization enhances the complexity of the entire system and costs a significant amount of processing time. Therefore, it is imperative to utilize computer-aided systems for these purposes.

The rapid progress in artificial intelligence (AI), especially in the field of deep learning (DL), presents a chance to completely transform this crucial area of healthcare [10, 11]. Deep learning, an area of artificial intelligence, utilizes artificial neural networks with numerous layers to effectively represent and comprehend intricate patterns. Within the realm of medical imaging, deep learning models can undergo training to achieve exceptional precision in identifying and categorizing various cell kinds, illnesses, or irregularities [12]. This skill's potential is highly significant in the field of hematology because it plays a critical role in accurately detecting and classifying blood cells, which is essential for diagnosing a variety of illnesses [9].

The strong point of this work is that we modify the Faster RCNN [13] for generic object identification and instance detection to specifically identify cells in blood images. The method's main advantage is that it improves detection accuracy while reducing training time. Faster RCNN was used as the base network to generate the feature map which will be fed to the Region Proposal Network (RPN) [14]. RPN is usually used to segment and detect WBCs in medical images, i.e. blood smear images, because of their capability to identify regions of interest with high speed and precision. Researchers have introduced different methods and techniques for the detection and classification of leukemia. While there are still some limitations in this field, the challenges present in current methodologies drive the motivation behind this study. The main contributions of this study are as follows:

1. We have developed a hybrid framework for extracting features from WBCs. This approach combines moment classification with CNN-based feature extraction, using feature fusion.
2. We modified the faster RCNN framework, which is used for generic object identification and instance segmentation to specifically identify cells in blood images. And used ResNet 50 for backbone model for best feature extraction with RPN Network
3. We show that training MCNet on a dataset with pixel-level annotations for several tasks not only enhances the identification of blood cells on this dataset, but also enables labelling of bold cell areas on an unlabeled dataset.
4. Fine tune YOLO v9 for show the comparative results with MASK RCNN and Faster RCNN

The remainder of the paper is organized as follows. Section 2 provides an overview of several related works. Section 3 describes the proposed model for detection Model for blood cells with MCNet. The dataset, experimental results, and analysis are presented in Section 4, and the proposed study is concluded in Section 5.

2. Literature review

Researchers have shown increasing interest in the use of DL in recent years [11]. Several application areas extensively utilize it to address segmentation, classification, and detection issues. In addition, hematology and blood smear analysis have been prominent areas of research, garnering interest from professionals in the medical field and the technology industry for many years. Researchers in the IT discipline have mostly concentrated on three specific areas in relation to the analysis of peripheral blood

smears: detecting malaria, diagnosing leukemia, and classifying blood cells [15].

The authors of this study [16] developed a classification method that uses a convolutional neural network (CCN) to effectively discriminate between eight unique cell types present in the circulation. The authors proposed a method of refining the training process for an all-inclusive classifier by leveraging a dataset of more than 17,000 cell images collected from clinical practice. The classifier uses a combination of Vgg-16 and Inceptionv3 models. the overall classification accuracy is at 96.2%. However, this method has longer trained times due to its depth and complexity, and classification accuracy is still low.

This study [17] presented a technique for automatically classifying blood cells. This method used advanced deep learning models, such as vision transformers with a 16-bit patch size (ViTb16) and pre-trained convolutional neural networks. It also used a custom multi-layer modified convolutional neural network model. Researchers have achieved impressive outcomes on the 11-class PBC dataset by employing transfer learning, fine-tuning, and ensemble learning techniques. The most remarkable achievement is the exceptionally optimized EfficientNetV2 B0 model, which demonstrated exceptional performance on the original PBC dataset. The model obtained a macro-average precision, recall, and F1-score of 91%, 90%, and 90%, respectively, as well as an impressive average accuracy of 93%.

In this article [18], The authors compared Google ViT and DL ImageNet CNNs, which could classify four kinds of white blood cells (WBCs) from two sets of peripheral blood smears. This PBC dataset comprises 17,092 superb images of eight distinct categories of blood cells. The use of the three balanced PBC datasets provides conclusive evidence of Google ViT's advantages and superior performance compared to ImageNet CNNs. Contrarily, the BCCD dataset consists of 349 low-quality images of four different types of white blood cells (WBCs). Three well-balanced datasets (DS-4, DS-5, and DS-6) accompany it, generated using data augmentation techniques. This research showcased the robustness, reliability, and resistance of Google ViT while dealing with noisy data, in contrast to ImageNet CNNs. This approach needs a substantial quantity of labeled training data to achieve optimal performance.

This study [19] demonstrated and applied the blood cell detection technique using the YOLOv5 (YOLOv5-ALT) model. The goal of this research was to improve detection precision using YOLO methods. This study introduces an improved technique that addresses the limitations of the previous approach. The author changed the bounding box regression loss function, added an attention mechanism to the feature channel, and tweaked the SPP module in the YOLOv5 backbone feature extraction network to make this happen. Additionally, the author assesses the model's efficacy by comparing each evaluation index using the deep learning object detection technique. This approach is more aligned with the efficacy of the blood cell detection job. However, YOLOv9 has a propensity for generating an increased number of false negatives. It adopts a more cautious detection methodology, which mitigates false positives but heightens the likelihood of overlooking genuine objects.

In this study [20], the researchers designed an object detector to identify various blood components, including white blood cells, red blood cells, and platelets. The detector, called FED (Fast and Efficient YOLOv3), is a one-stage detector similar to YOLOv3. It performs detection on three distinct scales. To improve efficiency and adaptability, the proposed object detector uses the Efficient Net Convolutional Neural Network at its core. Furthermore, the authors used dilated convolution to improve the backbone's receptive field. In addition, they employed the Depth wise Separable Convolution technique to reduce the detector's parameters. The training dataset, the BCCD dataset, yielded average accuracy values of 90.25% for platelets, 80.41% for red blood cells, and 98.92% for white blood cells. Nonetheless, YOLOv3 has difficulties recognizing tiny objects in images because of the configuration of its bounding box predictions.

The author presents a one-stage network that utilizes an enhanced version of YOLOv5 to accurately detect blood cells [21]. The first step is to integrate the transformer and bidirectional feature pyramid network (BiFPN) into both the backbone network and neck network. The purpose of this step is to improve and refine the adaptation. The outputs of the neck network also include the Convolutional Block Attention Module (CBAM), which improves important features in both space and across channels. Furthermore, introducing an Efficient Intersection over Union (EIoU) aims to improve the model's accuracy and performance in terms of localization. The improvements have been integrated into the YOLOv5s model, resulting in the creation of YOLOv5s-TRBC. The studies conducted on the blood cell dataset (BCCD) demonstrate that the proposed technique achieved a mean average accuracy (mAP) of 93.5% in detecting the three types of blood cells.

In this paper [22], the authors presented a novel neural network architecture that combines the features of convolutional neural networks, notably Xception, with recursive neural networks, specifically LSTM. As a result, the author used the combined Xception-LSTM framework to categorize blood cell pictures. This approach preserves both the chronological and geographical details of the visual input and is capable of extracting organized information from these components. The methodology differs from previous manual feature extraction methods in that it does not rely on cytoplasmic or nuclear segmentation. Alternatively, it has the capability to automatically extract and classify the underlying features hidden inside cell image patches. The author's proposed approach surpassed the previously established approaches in terms of classification accuracy when applied to the blood cell dataset. The classification accuracy is 90.79%. However, this framework is more resource-intensive and has low classification accuracy.

This study introduced WS-YOLO [23], which is light weight enhanced blood detection method. DWS-YOLO achieves a favorable trade-off between detection accuracy, computational complexity, and inference speed. We improve the YOLOv5-Nano's network architecture by adding a lightweight C3 module, a better attention mechanism, an effective loss function, and more soft non-maximum suppression. Thanks to the improvements, our detector shows significant promise in the realm of cell detection applications. The author suggested model has significantly improved several evaluation criteria, including the number of parameters, computational complexity, and detection accuracy. Experimental findings on various datasets have verified this improvement. Our model outperforms the most sophisticated object detection algorithms in terms of achieving state-of-the-art (SOTA) outcomes and demonstrating superior resilience performance on the BCCD dataset. Furthermore, it has exceptional generalization capabilities when used with the Raabin-WBC dataset. The information above clearly indicates that our suggested detector is a smaller, faster, and more precise blood cell detector. The average blood detection accuracy is 93.8. Nonetheless, YOLOv5 has difficulties recognizing tiny objects in images because of the configuration of its bounding box predictions.

In this paper [24], the authors introduced the Enhanced Channel Attention Module (ECAM), a method that utilizes an attention mechanism to enhance the accuracy of blood cell detection. When applied to the BCCD dataset, the method produced results comparable to those of other attention processes. They created Enhanced-CBAM by combining the spatial attention mechanisms in CBAM and ECAM. The author then utilized this to develop a new network known as the Enhanced Channel Attention Network (ENCANet). By comparing and testing many different methods, ENCANet clearly does a better job than other methods for finding blood cells with just 6.5 million parameters. Nevertheless, this methodology may lack efficacy in more intricate cases, such as images with excessively congested blood cells. The network achieved an accuracy of 90.3 AP with a parameter size of only 6.5 M, even though the classification accuracy is still low.

3. Methodology

First, we attempted to solve the problem of instance detection. Therefore, the application is part of the natural evolution of object detection technology. To shift from grainy box-level category recognition to precise pixel-level classification, we choose to use the convolutional neural network (CNN) for the purpose of feature extraction and classifying the blood cell. In order that, Deep learning enables the use of images as input, resulting in speedier analysis. Additionally, by considering that each pixel in a medical image has valuable information, the risk of data loss due to feature extraction may be avoided.

The ResNet50 [25] model designs its input layer to handle data from a dataset with dimensions of 224x224. Afterwards, the convolution layer, which follows the input layer, modifies its values. The hybrid model modifies layers with numbers 177 and above. More specifically, we eliminated the previous five levels and replaced them with 10 new tiers. As a result, the model's overall layer count increased from 177 to 182. Furthermore, the new model eliminates the input, convolution, activation, pooling, fully connected, SoftMax, and classification layers from the ResNet50 model, forming its foundation. We append two distinct layers, completely interconnected, to this foundation. Neural networks employ batch normalization as a technique to standardize the input values of each layer, leading to enhanced stability and accelerated model performance. Normalizing the data dimensions from other levels is advantageous. This not only improves the design's efficiency, but also adjusts the size and plots the input data within a specific range. The process of normalization may be represented by Eq. (1) and (2).

$$x^n = \frac{x^n - E[x^n]}{\sqrt{Ave(x^n) + \varepsilon}} \quad (1)$$

x^n : represent the dimension of the first input layer. $E[x^n]$: represent the function of average dimension.

The standard deviation adds to $\sqrt{Ave(x^n) + \varepsilon}$. The primary objectives of pooling are to reduce the input size of the data for the next layer. The structures on this layer do not include any learning process. The primary purpose of this layer is to minimize computing complexity. Commonly used techniques include Average Pooling and Maximum Pooling.

$$S = w * h * d * 2 \quad (2)$$

were,

$w = \frac{w-f}{A+1}, h = \frac{h-f}{A+1}$

w : width of input image, h : high of input image, d : depth of input image, f : filter size. A : steps. S : size of generated information.

The SoftMax layer receives the data from the preceding layer and generates the probabilistic value as part of the classification process. The SoftMax layer gives a value to indicate the class to which it is most similar while doing classification. The deep learning network utilizes probabilistic calculations to determine the value produced in each layer. Eq. (3) and (4) represent the process of SoftMax Layer and probability consecutively.

$$p[y = jx^n, w, b) = \frac{exp^{x^n wj}}{\sum_{j=1}^{n} exp^{x^n wj}} \quad (3)$$

$$\sigma[y = jx^n, w, b) = \frac{e^{x^n wj}}{\sum_{j=1}^{n} e^{x^n wj}} \quad (4)$$

were,

p : denote probability.

σ : denote SoftMax layer.

x: is main classes.

W, b: a weight vector.

This section will analyze the result of the instance segment carried out by the Mask R-CNN model and the classification results obtained using the majority vote procedure. We will evaluate it using both quantitative and qualitative measures.

The first phase involves the use of the RPN. It suggests potential boundary boxes for objects without considering their specific classifications. The second step is referred to as the R-CNN stage. In this stage, features are extracted using RoIAlign for each proposal. The stage also involves performing proposal classification, bounding box regression, and mask prediction. In the first level, a Faster R-CNN enables the detection and extraction of individual white blood cells. This model is an object identification method that enhances Fast R-CNN by including an RPN with the Resnet model. In the Faster R-CNN framework, the Region Proposal Network uses an image to produce a collection of rectangular objects. Each item is assigned a score that represents its likelihood of being an actual object. These suggestions are generated by moving a tiny network over the feature map produced by the last shared convolutional layer, using a n x n spatial window that is mapped to a feature with lower dimensions.

To build a feature map while the procedure is being carried out, the whole picture is first processed by using several convolutional and maximum clustering layers. A region of interest (RoI) clustering layer is charged with the responsibility of extracting a feature vector of a specified length from the feature map for each individual item proposal in the subsequent phase. The input of each feature vector is then sent into a sequence of completely linked layers, which ultimately split out into two output layers that are sister to one another.

As opposed to the other layer, which provides a general "background" type, one of these layers generates SoftMax probability estimates over N object types. One further layer is responsible for the generation of four real-valued numbers for each of the N object classes that the objects belong to the SoftMax layer takes all input vector from (n_1 ……., n_k) then apply normalize into probability for n classes.

Fig.1 illustrates the connection between the final convolution feature map of the Convolution Network and a sliding window (SW), as indicated by the yellow arrows in the diagram. The Region Proposal Network (RPN) has two output layers that yield the categorization scores and bounding boxes (Bbox) for the suggested regions. The red arrows demonstrate the salient characteristic of the Fast R-CNN Network. We link the last convolutional feature map to a Region of Interest (RoI) Pooling Layer. The term "region of interest" (RoI) denotes the designated area under consideration.

However, a residual network ResNet50 and a region proposal network RPN serve as the foundation for this architecture.

The ResNet-RPN can produce multi-scale feature maps since it only requires a single-scale input image to be processed.

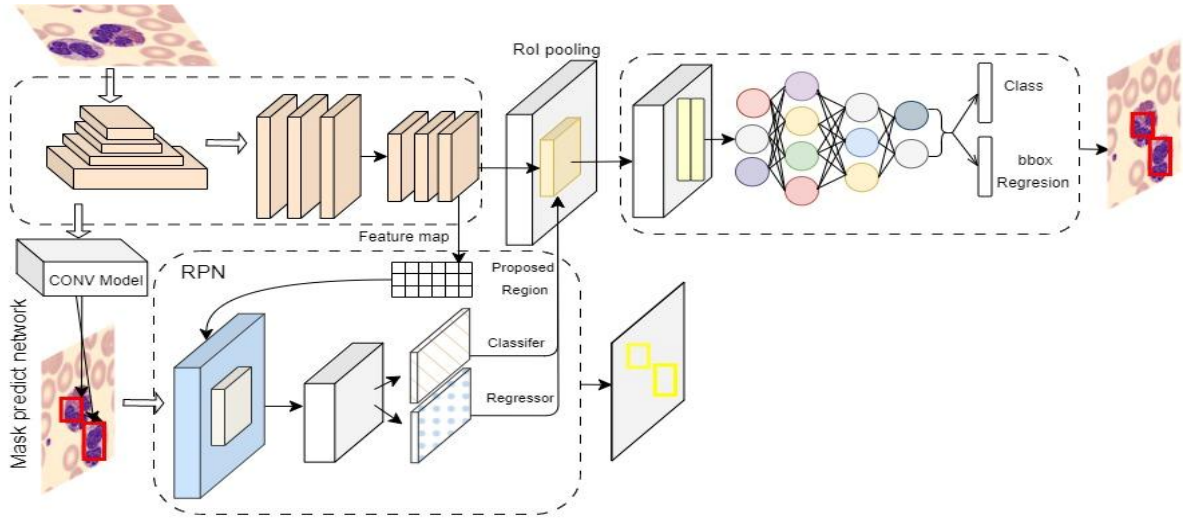

Figure. 1 The suggested cancer cell detector utilizes the MCNet architecture

The RPN [26] is a method that generates potential object-like regions by systematically moving a predetermined set of filters with a single receptive field over the highest-level convolutional feature maps. For instance, the region proposal network (RPN) plays a key role in generation of full various sizes and different aspect ratios. This infrastructure is built out from the sliding subnetwork at every point from the multi-level feature maps. The anchor referred to box of dimension and abrasive is presentation, and it is used to distinguish every plan proposal. The keys in the features pyramid represent different levels which are basically discussed by anchor scales. To capture all cells, regardless of their size or form, our method adheres to the default arrangement of RPN anchors. An application of a deep detector for cancer cells is presented in this work, which is built on the framework of Faster R-CNN. This detector is intended to be used in medical image processing. And then, taking into consideration the data shown above. A network that traverses the RoIAlign layer to extract a vector of features for each item proposition are the bounding box recognition and mask prediction network. The features patch of each proposal is taken from the features pyramid at a certain level, with the level being chosen based on the size of the proposal.

Following this, the characteristics of the proposal are input into two subnetworks. These subnetworks are the mask prediction network and the object classification and bounding box regression network. When compared to a ground truth box, an item proposal with an Intersection over Union (IoU) value equal to or greater than 0.5 receives a foreground label. We do this to assist in training the bounding box recognition and mask prediction networks. Meanwhile, these prominent concepts are accountable for sampling 33 percent of the total number of picture areas of interest.

3.1 Using mcnet for blood cell detection

The convolutional feature extractor may accept an image of any dimension as input and generate many hierarchical features as output. The feature extractor's design is critical because it directly impacts the detector's speed, memory use, and overall performance. The number of parameters and the types of layers employed determine this. Cell-pixel-level annotations train the entire MCNet model, including both detection and semantic segmentation networks, but the detection phase only predicts the bounding box locations and classification probabilities. Disregarding the mask branch can reduce the inference time without affecting the detection result. The masking procedure involves creating a mask and labeling a bounding box for each cell's ground truth. During the detection procedure, we exclude any regions without cells and without annotated centroids from further consideration. The lack of adherence to industry standards in histology facilities regarding the staining and acquisition procedures accentuates the variation in hue observed in histological pictures. Prior to calculating the cell mask and bounding box labels, we normalize the blood cell images using the stain normalization technique described in [3]. This enables us to overcome the challenge that we are now encountering.

3.2 Estimate cells mask and bounding box labels

In the case of blood cell datasets, we do not have individual pixel-level annotations to rely on and only those images which are classified with each class are

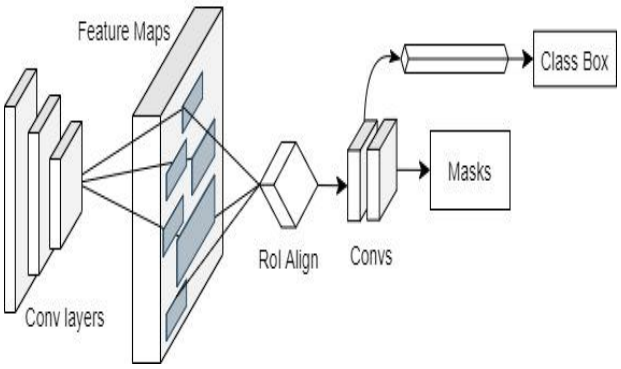

Figure. 2 MCNet framework to generate the cell mask and bounding box (bb) for unlabeled dataset

provided to us. However, even though the results of detections in most cases will be insufficient to L train our model within cell detection and segmentation, we need a ground truth for instances and cell masks ground truth that is an extremely difficult and time-consuming task. This vision holds such because the imagery present contains more than 17000 images. We add automatic annotation to MCNet model to generate masks for all images.

3.3 Faster RCNN

Fig. 2 illustrates the object identification framework we have developed. The four main components of Faster R-CNN are as follows: To begin with, let's discuss convolutional layers. The Faster R-CNN [5] begins the process by extracting the feature maps of the picture using a series of fundamental convolutional, ReLU, and pooling layers. These maps then find their way to the RPN and fully linked layers below. Furthermore, there are regional proposal networks. We utilize the RPN network to generate regional proposals. The layer uses the SoftMax function to categorize whether the anchors belong to the foreground or background. Afterwards, it uses bounding box regression to correct the anchors and provide accurate suggestions. Additionally, ROI pooling is a computer vision approach employed to extract distinctive characteristics from specific areas of interest. This layer collects the input feature maps and recommendations. Following the synthesis of the input, we extract the feature maps and forward them to the next fully connected layer for object categorization. Furthermore, categorization, we use the bounding box regression technique to precisely determine the final position of the detection frame. Furthermore, the proposal feature maps are used to determine the proposal's category. These two operations are performed together. Faster R-CNN is a competitive option in the field of region-based

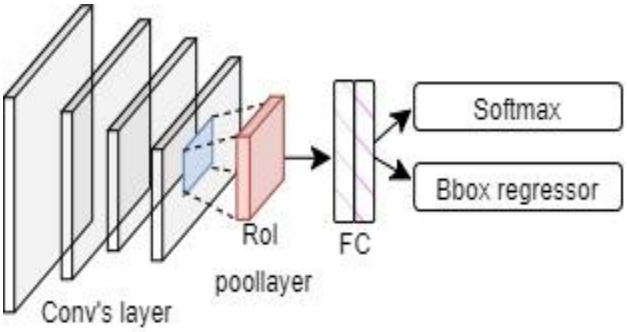

Figure. 3 Representation of Faster R-CNN segmentation for our proposed model

detection networks, as seen in Fig. 3. To estimate blood cell masks and bounding box labels for the unlabeled dataset, MCNet was used.

To construct the cell mask and bounding box labels for the unlabeled dataset, every segmented blob that was predicted by MCNet is used. Following the completion of many convolutions, it creates a single feature map, then generates proposals by means of the RPN network, and finally, it executes object detection on the last layer of the feature map once the ROI pooling process has been completed.

During training phase, we determine the parameters as W of the proposal network label, arned from a set of training samples $T_s = [X_n, Y_n]_{n=1}^{N}$, where *Xn* is a training image patch, and *Yi* = (*yi*, *bi*) the combination of its class label, $yi \in \{0, 1, 2, \cdots, K\}$ and bounding box coordinates *bi* = (*bxi* , *by i* , *bwi* , *bhi*). This is achieved with a multitask loss function. Eq. (5), (6) and (7) represent the loss functions.

$$lw = \sum_{N=1}^{N} \sum_{i \in s^n} \alpha_n l^n [X_i, Y_i, w] \quad (5)$$

Lw: total loss function

αn: weight coefficient that determines how much each task or dataset contributes to the overall loss.

The variable N represents the number of detection branches. The weight of the loss is represented by the variable m. S is a set consisting of S1, S2…, SM, where *Sm* holds the samples of scale m. Understand that scale only influences the selection of a specific subset of training samples, denoted as *Sm*, and this subset is the only one that impacts the loss of detection layer m. The loss function of each detection layer integrates these two objectives.

$$l(X, Y|w) = -log\, p_y(x) + \lambda(y \geq 1)\, log(R_b, \bar{R_b}) \quad (6)$$

Where, p(*X*) = (*p*0(*X*), · · ·, *pK*(*X*)) is the probability distribution for classes, *λ:* denote a trade-off coefficient, −log *py*(*X*) denote the cross-entropy

loss and $\bar{R}_b$ denote the regressed bounding box. The bounding box loss is only used for positive samples and the optimal parameters **W**.

3.4 Object detection network

While the proposal network has the potential to function as a detector, its effectiveness is limited because of its inadequate coverage of objects with its sliding windows. An additional detecting network has been included. The first step involves using a ROI pooling layer to extract features with a predetermined dimension. The characteristics are then input into a fully linked layer and output layers.

$$\ell(w, w_D) = \sum_{n=1}^{N} \sum_{i \in s^n} \alpha_n * l^n(X, Y|w) \qquad (7)$$

l(w, wD): represents the total loss function.
$\alpha_{n:}$ weight.

The variables "n" and "n+1" denote the loss and training samples for the detection subnetwork. The classification and a smoothed loss for bounding box regression. The detection sub-network utilizes a subset of the proposal sub-network parameters W and introduces supplementary parameters WD. approach utilizes a cross-entropy loss for the parameters to be adjusted in unison. The recommended solution applies ROI pooling to the "conv4-3" layer instead of the "conv5-3" layer. Our research bases the decision on the exceptional performance of the "conv4-3" feature maps. One possible theory is that the feature map "conv4-3" exhibits a higher level of complexity and is better suited for precise prediction of bounding box placement.

4. Result and discussion

4.1 The experiment setting

We assigned the values for momentum and weight decay to 0.9 and 0.0001, respectively. At the outset, we train the network heads for 50,000 iterations with a learning rate of 0.001. Afterwards, we proceed to train the layers beginning from stage 4 of ResNet-50 using the identical learning rate for a total of 80,000 iterations. Finally, we adjust all of the network layers using a learning rate of 0.0001 for a total of 40,000 iterations.

4.2 Dataset

The classification of blood cells was performed using the peripheral blood cells dataset (PBC) [27], which consists of 17,092 RGB image of normal blood cells. These images were obtained using the CellaVision DM96 analyzer. The image has a dimension of 360x363 each of them which has been labelled by experienced pathologists of Clinical Hospital Barcelona. Privacy of individuals was maintained by removing their connections with the data as well as the source, thus creating a set without personal records yet with the same level of accuracy for scientific purposes. What we would like to stress is the fact that the subjects were chosen from those individuals who had not been previously diagnosed with any infections, blood disorders, cancer and those who were on any medication at the time of cheque. We divided the dataset, keeping 80% of the samples in each class for training, 20% for testing. We ensured that the ratio of samples per class in all sets remained the same as in the original dataset.

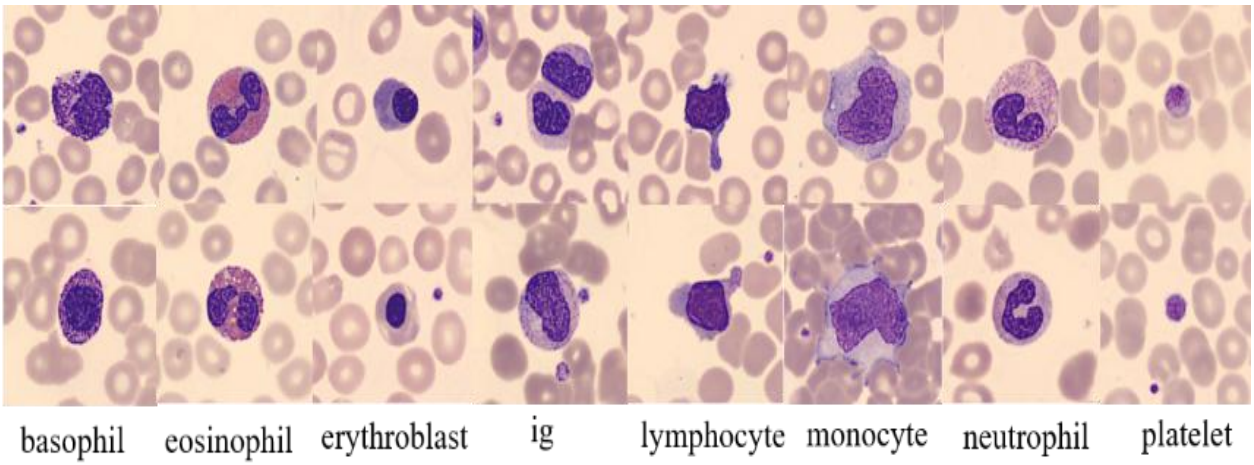

Figure. 4 displays representative examples from 8 distinct classes of the PBC dataset: (a) These classes include Basophils, (b) Eosinophils, (c) Erythroblasts, (d) Lymphocytes, (e) Monocytes, (f - h) IGs, (i) Neutrophils, and (j) Platelets. IG1, IG2, and IG3 correspond to the metamyelocytes, myelocytes, and promyelocytes, respectively.

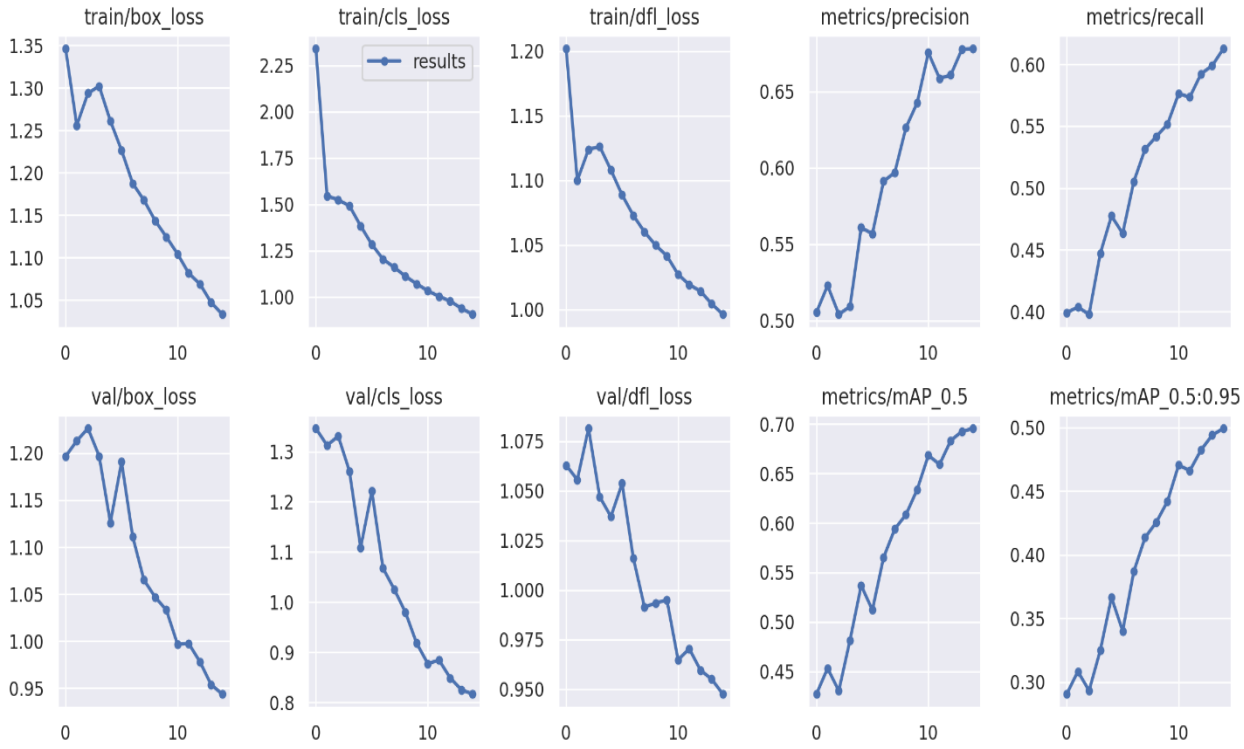

Figure. 5 Metrics of YOLOv9 as baseline performance of detection

The PBC dataset has a total of eight distinct cell types, including neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (IGs), erythroblasts, and platelets. Fig. 4 presents exemplary instances from eight different classes of the PBC dataset. In addition, we used a second type of dataset named BCCD. It is publicly available dataset of annotated blood cell images called Blood Cell Count Dataset (BCCD)[28].

4.3 Fine tune yolov9 on blood cell

YOLOv9 is the most recent version in the YOLO (You Only Look Once) family of real-time systems used for detecting objects.

This version of the software improves upon its predecessors by integrating state-of-the-art deep learning algorithms and architectural design, resulting in exceptional performance in tasks related to object identification.

YOLOv9 enhances the previous version, YOLOv7, by using the Generalized ELAN (GELAN) architecture and Programmable Gradient Information (PGI) to broaden its functionalities. This firmly establishes itself as the foremost real-time object detector of the present age. Because of the considerable impact that the YOLO model has had on the area of computer vision, researchers have been motivated to continually enhance and broaden the capabilities of the model. After increasing both the size of the image and the pace at which it is learned, it has been discovered that there is an improvement. In comparison to the results of the other studies, this one was able to produce a mAP50 value of 0.953, which is much higher. The hyper-parameter setting is tune for lr = 0.001, lrf = 0.01, momentum = 0,937, weight decay = 0.0005, warmup_epochs=3.0, warmup_momentum = 0.8, warmup_bias_lr =0.1, box=7.5 optimizer: SGD (lr=0.01), 0.785 mAP50. Fig. 5 illustrates the training and validation results metrics of YOLO v9. On the other hand, Loss of training and validity show in the same figure which is decreasing gradually to improve the performance.

Fig. 6, and 7 show the confusion matrix. A matrix is a tabular representation that provides a concise summary of a classification model's performance by comparing the predicted labels with the actual labels. The output shows the count of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) resulting from the model's predictions. A precision-recall curve illustrates the relationship between precision and recall, showing that precision often decreases as recall increases. Alternatively, one can compare data for a specific level of one measure with another (e.g., precision at a recall level of 0.75), or combine both measures into a single measure. The effectiveness of our proposed model in terms of precision and recall is shown in Fig. 8 illustrate the precision recall cure of MaskRNN model.

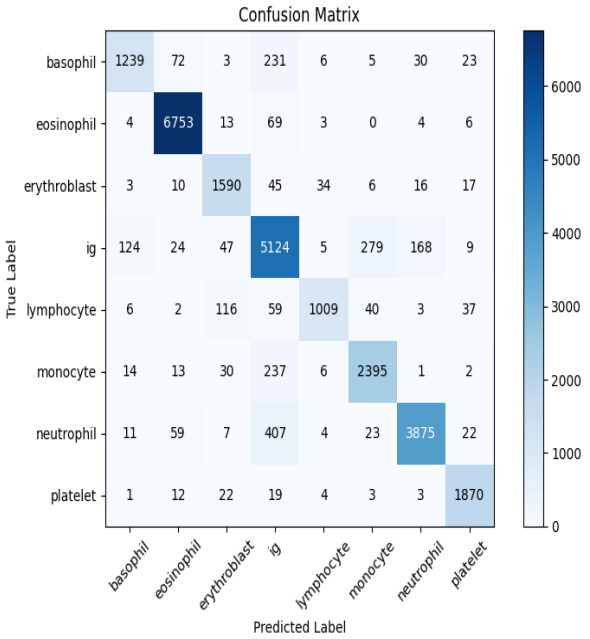

Figure. 6 confusion matrix of Mack RCNN

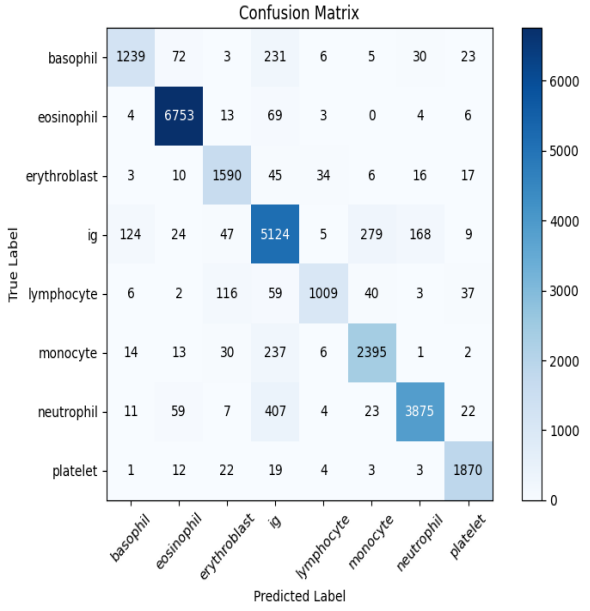

Figure. 7 confusion matrix for MCNet model

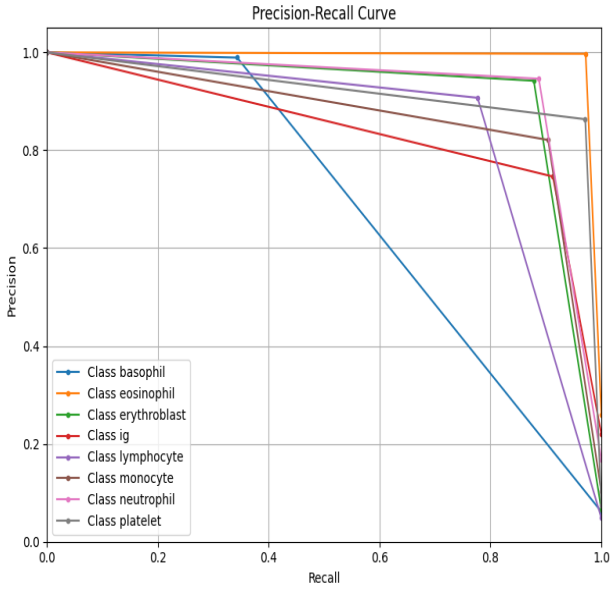

Figure. 8 Precision – recall Curve of Mask RCNN

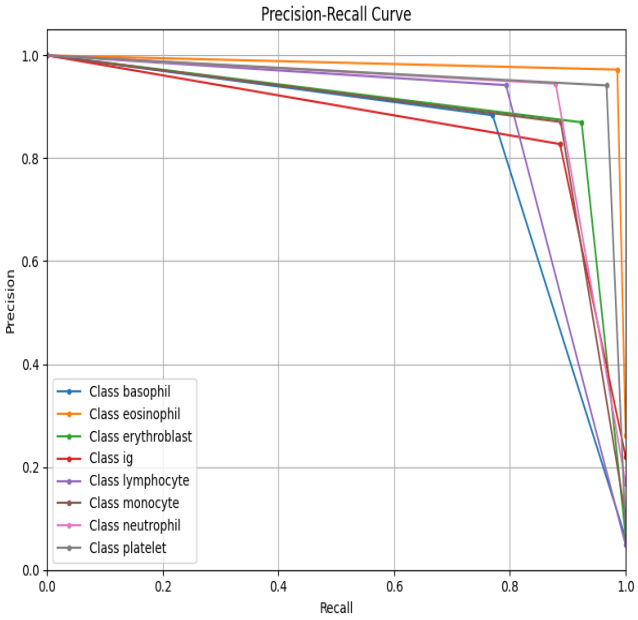

Figure .9 Precision – recall Curve of MCNet

The PR curve analysis in Fig.8 indicates that the model has strong performance for several classes (e.g., eosinophil and erythroblast), however encounters accuracy challenges for others, particularly when recall rises. Mitigating class imbalance and enhancing the model may result in improved performance across all categories.

Fig. 9 shows a Precision-Recall Curve for a multi-class classification of our proposed model MCNet, illustrating many classes represented by distinct colored lines. Each line illustrates the trade-off between accuracy and recall for a particular class.

Fig. 10, 11, and 12 the detection results of using three models (YOLOv9, MASK RCNN, and MCNet). They show that MCNet is more accurate than the others. To show the effectiveness of our model, we utilized another dataset named Blood Cell Count Dataset (BCCD).

Fig. 13 and 14 illustrate the average precision-recall curve that reveals the effectiveness of our proposed model MCNet. Fig. 15 shows the detection results when running MCNet on the BCCD dataset. Fig 16 illustrates ROC curve of our proposed model MCNet for BCCD dataset. The figure shows the AUC results related platelets, RBC and WBC.

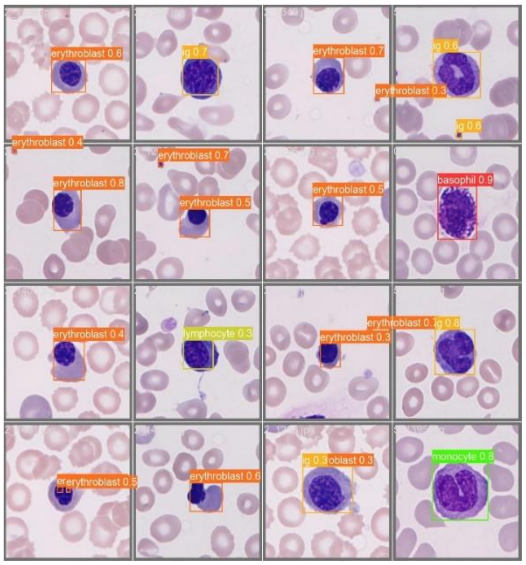

Figure. 10 depicts the results of detection for YOLOv9

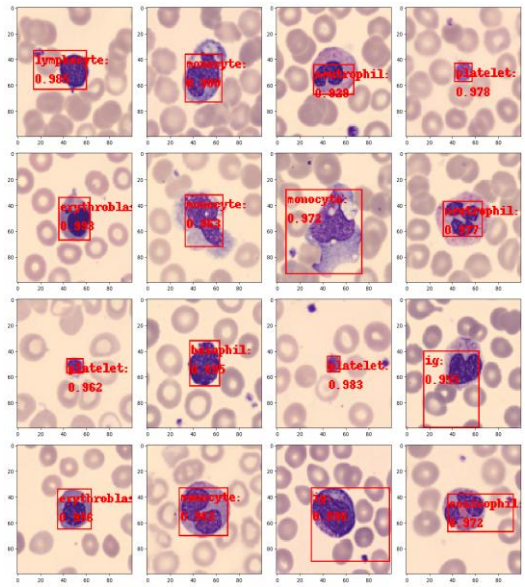

Figure. 11 the results of detection for Mask RCNN

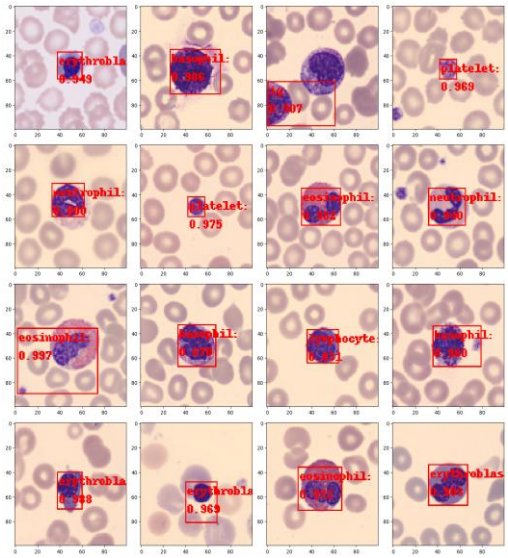

Figure. 12 The detection results of MCNet model

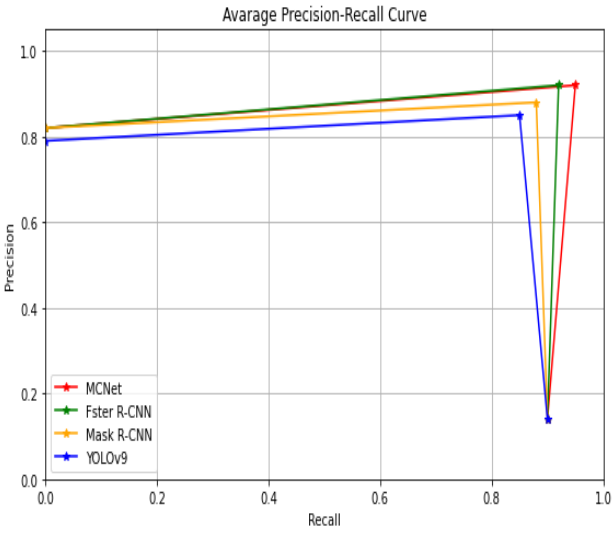

Figure. 13 comparison precision- recall curve of MCNet, Mask RCNN, and YOLOv9 using BCCD dataset

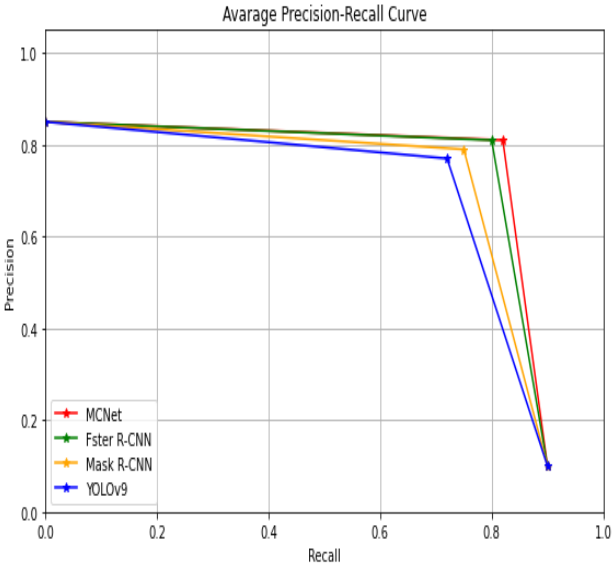

Figure. 14 comparison precision- recall curve of MCNet, Mask RCNN, and YOLOv9 using PBC dataset

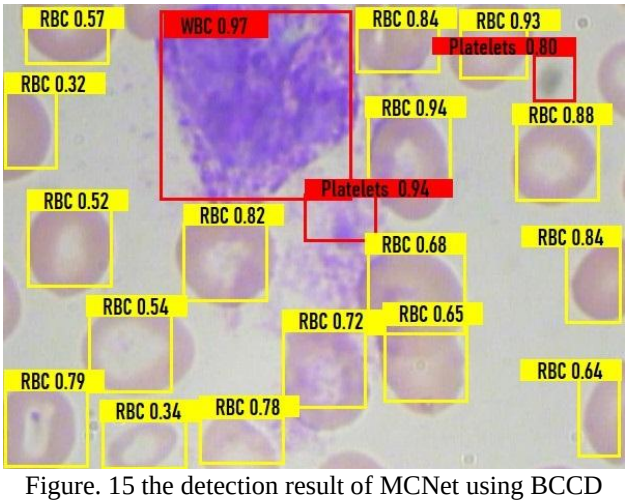

Figure. 15 the detection result of MCNet using BCCD dataset

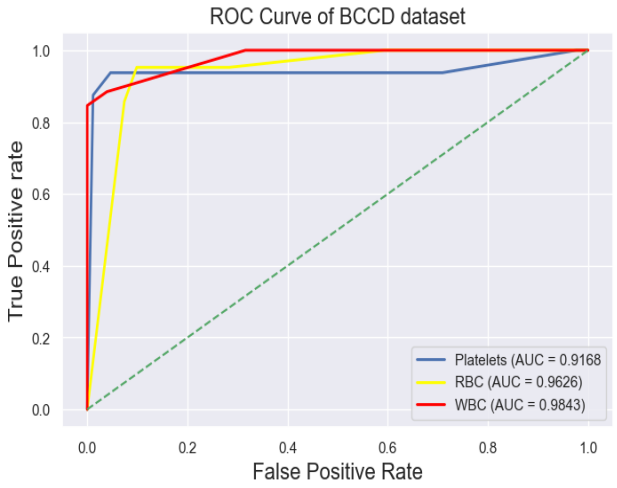

Figure. 16 ROC curve of our proposed model MCNet for BCCD dataset. The figure shows the AUC results related platelets, RBC and WBC.

Table 1. comparison of three implement models on PBC dataset

Method	Recall	Precision	mAP@IoU0.50	mAP
YOLOv9 (ours)	73.46	77.86	94.35	0.785
MASK RCNN (ours)	75.38	80.34	93.22	78.53
Faster RCNN (ours)	79.83	81.41	94.46	0.815
MCNet (ours)	81.23	81.72	95.70	0.82

Table 2. Comparison of proposed model on BCCD dataset

Method	**Recall %**	**Precision %**	**mAp@IoU0.50%**	**mAP**
YOLOv9	88.46	87.86	85.45	88.36
MASK RCNN	90.38	89.34	88.36	86.53
Faster RCNN	93.83	92.41	93.98	91.56
YOLOv5s-TRBC [5]	88.80	87.40	93.50	*
YOLOv5 [5]	87.70	87.90	92.20	*
AYOLOv5 [29]	91.50	86.20	89.90	*
YOLO5-csw [30]	85.70	84	89	*
MCNet (proposed)	**97.23**	**96.72**	**96.76**	**95.76**

Table 1 describe that MCNet model outperforms the other models in all metrics for PBC dataset.

Table 2 describes that MCNet model outperforms the other models in all metrics for BCCD dataset, demonstrating the highest recall, precision, mAP@IoU 0.50, and mAP. It is probable that the model is more sophisticated, with enhanced object localization and feature extraction capabilities. In the majority of metrics, Faster R-CNN outperforms YOLOv9 and Mask R-CNN, particularly in terms of mAP, and follows MCNet. Mask R-CNN and YOLOv9 exhibit comparable performance; however, Mask R-CNN exhibits a minor advantage in overall mAP@IoU 0.50, as YOLOv9 has inferior precision and recall.

5. Conclusion

In this context, we introduce the Mask Cell of the multi-class deep network (MCNet) as an innovative architecture designed to enhance the learning quality of predicted instance masks. Our concept entails using a faster R-CNN to segment white blood cell microscopic images, enabling precise classification of cases of acute lymphoblastic leukemia. This method seeks to improve the efficiency and efficacy of the diagnostic procedure. The proposed network block evaluates the proposed mask, Intersection over Union (IoU), by integrating the instance feature with the corresponding expected mask. This study employed a transfer learning methodology to implement Mask R-CNN for the segmentation of white blood cells in microscopic images. To rectify the problem of inadequate illumination in stained white blood cell microscopy images, we incorporated a contrast enhancement process into the image dataset. The comparison experiment utilizes YOLO version 9 for classification and Mask R-CNN. The MCNet methodology calibrates the disparity between mask quality and suggested detection; hence, it augments the efficacy of instance segmentation. The conclusive outcomes for the two datasets trained using PBC and BCCD are as follows: The mAP@IoU0.50 accuracy for the PBC dataset is 95.70, but the accuracy for the BCCD dataset is 96.76, with recall and precision recorded at 97.23 and 96.72, respectively. In our future work, we will further investigate other deep learning algorithms to participate in the development of healthcare models.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

For this research work all authors have equally contributed to Conceptualization, methodology, validation, resources, writing—original draft preparation, writing—review and editing.

References

[1] M. Li *et al.*, "A deep learning model for detection of leukocytes under various interference factors", *Sci Rep*, Vol. 13, No. 1, pp. 2160, 2023

[2] M. Sebai, X. Wang, and T. Wang, "MaskMitosis "a deep learning framework for fully supervised, weakly supervised, and unsupervised mitosis detection in histopathology images", *Med Biol Eng Comput*, Vol. 58, No. 7, pp. 1603-1623, 2020.

[3] I. S. Rahat, M. A. Ahmed, D. Rohini, A. Manjula, H. Ghosh, and A. Sobur, "A Step Towards Automated Haematology: DL Models for Blood Cell Detection and Classification", *EAI Endorsed Trans Pervasive Health Technol*, Vol. 10, 2024.

[4] Z. Khan *et al.*, "A Framework for Segmentation and Classification of Blood Cells Using Generative Adversarial Networks", *IEEE Access*, Vol. 12, pp. 51995-52015, 2024.

[5] Y. He, "Automatic Blood Cell Detection Based on Advanced YOLOv5s Network", *IEEE Access*, Vol. 12, pp. 17639-17650, 2024.

[6] H. Song and Z. Wang, "Automatic Classification of White Blood Cells Using a Semi-Supervised Convolutional Neural Network", *IEEE Access*, Vol. 12, pp. 44972-44983, 2024.

[7] A. Shakarami, M. B. Menhaj, A. Mahdavi-Hormat, and H. Tarrah, "A fast and yet efficient YOLOv3 for blood cell detection", *Biomed Signal Process Control*, Vol. 66, pp102495, 2021.

[8] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining Convolutional Neural Network with Recursive Neural Network for Blood Cell Image Classification", *IEEE Access*, Vol. 6, pp. 36188-36197, 2018.

[9] O. Islam, M. Assaduzzaman, and M. Z. Hasan, "An explainable AI-based blood cell classification using optimized convolutional neural network", *J Pathol Inform*, Vol. 15, p. 100389, 2024.

[10] H. K. Abass, M. H. Al-hashimi, A. S. Al-Zubaidi, M. Al-Mukhtar, "GastroFPN: Advanced Deep Segmentation Model for Gastrointestinal Disease with Enhanced Feature Pyramid Network Decoder", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 4, pp. 438-450, 2024, doi: 10.22266/ijies2024.0831.34.

[11] H. M. Al-Dabbas and M. S. Mahdi, "Classification of Brain Tumor Diseases Using Data Augmentation and Transfer Learning", *Iraqi Journal of Science*, Vol. 65, No. 4, pp. 2275-2286, 2024.

[12] A. S. Al-Zubaidi, M. Al-Mukhtar, and A. A. A. Baghdadi, "ALL-FABNET: Acute Lymphocytic Leukemia Segmentation Using a Flipping Attention Block Decoder-Encoder Network", *Journal of Image and Graphics*, Vol. 12, No. 3, pp. 239-249, 2024.

[13] R. Girshick, "Fast R-CNN." [Online]. Available: https://github.com/rbgirshick/

[14] S. Ren, K. He, R. Girshick, 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.1137-49, 2016.

[15] R. Al-qudah and C. Y. Suen, "Improving blood cells classification in peripheral blood smears using enhanced incremental training", *Computers in Biology and Medicine*, Vol. 131, pp.104265, 2021.

[16] A. Acevedo, S. Alférez, A. Merino, L. Puigví, and J. Rodellar, "Recognition of peripheral blood cell images using convolutional neural networks", *Comput Methods Programs Biomed*, Vol. 180, pp. 105020, 2019.

[17] M. A. Ali, F. Dornaika, and I. Arganda-Carreras, "Blood Cell Revolution: Unveiling 11 Distinct Types with 'Naturalize' Augmentation", *Algorithms*, Vol. 16, No. 12, pp.562, 2023.

[18] M. A. Ali, F. Dornaika, and I. Arganda-Carreras, "White Blood Cell Classification: Convolutional Neural Network (CNN) and Vision Transformer (ViT) under Medical Microscope", *Algorithms*, Vol.16, No. 11, pp. 525, 2023.

[19] Y. Guo and M. Zhang, "Blood Cell Detection Method Based on Improved YOLOv5", *IEEE Access*, Vol. 11, pp. 67987, 2023.

[20] A. Shakarami, M. B. Menhaj, A. Mahdavi-Hormat, and H. Tarrah, "A fast and yet efficient YOLOv3 for blood cell detection", *Biomed Signal Process Control*, Vol. 66, 2021.

[21] Y. He, "Automatic Blood Cell Detection Based on Advanced YOLOv5s Network", *IEEE Access*, Vol. 12, pp. 17639-17650, 2024.

[22] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining Convolutional Neural Network with Recursive Neural Network for Blood Cell

Image Classification", *IEEE Access*, Vol. 6, pp. 36188-36197, 2018.

[23] Y. Mao, H. Zhang, W. Wu, X. Gao, Z. Lin, and J. Lin, "DWS-YOLO: A Lightweight Detector for Blood Cell Detection", *Applied Artificial Intelligence*, Vol. 38, No. 1, pp. 2318673, 2024.

[24] F. Xu, H. Lyu, and W. Xiang, "Rethinking the Dilated Encoder in TE-YOLOF: An Approach Based on Attention Mechanism to Improve Performance for Blood Cell Detection", *International Journal of Molecular Sciences*, Vol. 23, No. 21, pp. 13355, 2022.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", In: *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[26] 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.1137-49, 2016.

[27] A. Acevedo, A. Merino, S. Alférez, Á. Molina, L. Boldú, and J. Rodellar, "A dataset of microscopic peripheral blood cell images for development of automatic recognition systems", *Data Brief*, Vol. 30, No. article 105474. 2020.

[28] M. M. Alam and M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells", *Healthcare Technology Letters*, Vol. 6, No.4, pp.103-108, 2019.

[29] W. Gu, K. Sun., "AYOLOv5 Improved YOLOv5 based on attention mechanism for blood cell detection", *Biomedical Signal Processing and Control.*, Vol.88 pp.105034, 2024.

[30] M. Huang, B. Wang, J. Wan, and C. Zhou, "Improved blood cell detection method based on YOLOv5 algorithm", In: *Proc. of 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* Vol. 6, pp. 992-996, 2023.