SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Lightweight White Blood Cells Detection Using Fusion of YOLOv5 and Attention Model

Nurasyeera Rohaziat[1,*], Mohd Razali Md Tomari[2], Wan Nurshazwani Wan Zakaria[3], Dipankar Das[4]

[1] Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[2] Department of Electronic Engineering, Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[3] School of Mechanical Engineering, College of Engineering, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia
[4] Department of Information and Communication Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh

ARTICLE INFO

Article history:
Received 18 October 2023
Received in revised form 16 April 2024
Accepted 6 June 2024
Available online 5 July 2024

Keywords:
White blood cells; WBCs; deep learning; detection; YOLOv5, eosinophil; lymphocyte; monocyte; neutrophil; attention module; squeeze excitation; convolutional block attention module

ABSTRACT

The human body is protected by an immune system which mainly consists of white blood cells (WBCs). There are five types of white blood cells, and each type will fight certain viruses and bacteria that are encountered in the human body. This defence system helps to maintain human health. Consequently, healthy WBCs keep humans healthy. Abnormality in WBCs can cause harmful viruses or bacterial infections. Leukaemia is a common WBCs disease which affects the production of good cells. Early detection is important for advanced treatment for cancer patient. One of the detection methods is by visual detection of the blood microscopic image since the five types of the WBCs are visually distinctive. In current practice, the pathologist will perform the diagnosis manually which may take time if there are many samples to examine. This procedure can be improved by automating it using a computer aided detection system. This paper studied the deep learning detection model of YOLOv5s and the effect of fusing the Squeeze-Excitation (SE) and Convolutional Block Attention Model (CBAM) into the YOLOv5s. It was performed on the four types of the WBCs, eosinophil, lymphocyte, monocyte, and the neutrophil taken from a public dataset. Based on the findings, the proposed method of YOLOv5s-SE, YOLOv5s-CBAM, and YOLOv5s-SE-CBAM produced overall accuracy of 99.5%, 99.5% and 99.4% mAP value and the performance are at par with the deeper model YOLOv5m with 65.8% of a smaller number of hyperparameters.

1. Introduction

The human immune system is a significant system that provides protection against infectious causing organisms such as bacteria, and viruses [1]. The immune system organ is called the lymphoid organs, which are located throughout the human body. The lymphocytes white blood cell habits in these organs. The white blood cells that are also known as the immune cells, are the ace in the system. Thus, maintaining healthy white blood cells is important for protecting the body against

* *Corresponding author.*
E-mail address: ge190052@student.uthm.edu.my

https://doi.org/10.37934/araset.48.1.117136

diseases. Any abnormality in the immune cells will affect human health and they can easily be harmed by infectious organisms. They may suffer from blood cancer like leukemia and other types of immune diseases [2]. Thus, it is important to detect any abnormality in an earlier stage so that it will be able to start an early treatment. In the present day, white blood cells are diagnosed by blood lab test, bone marrow test, and the cytochemistry test [3]. These tests usually will take time and tedious work. An automated diagnose method would be an advantage for both patient and pathologist. It will speed up the blood count test, white blood cells type classification and detection process.

Medical imaging technology has been developing over the years. The artificial neural network application in medical science has evolved since 2015 according to the study by Parveen *et al.,* [4]. Deep learning detection and classification in medical imaging is one of the fastest-growing methods in trend [5]. One of the CNN based medical imaging classifications is the MRI brain image that implied modified VGG16 and transfer learning [6]. The deep learning detection method can be divided into two categories which are the one-stage detection and the two-stage detection method. One-stage detection combining the feature maps and classification process. Whilst the two-stage detection method has two step processes. First is the searching feature maps process, then only the classification step. The one-stage detection method has been studied to be the fastest detection method [7]. You Only Look Once (YOLO) is one of the pioneers in the one-stage detection method [8]. The YOLO series keeps developing in terms of achieving a more robust and accurate detection model. The second YOLO version (YOLO9000) is designed not only to be faster and accurate, but also able to handle a larger number of classes [9]. In YOLOv2, the use of anchor boxes is improved. To determine the final prediction, it used the combination of anchor boxes and predicted offset. The use of batch normalization and the new loss function also improved accuracy in YOLOv2. Next is YOLOv3, which is claimed to be more accurate and faster than the previous version [10]. A convolutional structure, Darknet-53 was used in the model. Another improvement is the different scales and aspect ratios of the anchor boxes, which make it able to detect objects of different sizes and shapes. It also introduced a feature pyramid network (FPN) catering to different scale sizes. On the other hand, the YOLOv4 was developed by other researchers [11]. It employed the cross-stage partial network (CSPNet) for improvements. Then, YOLOv5 was developed by the Ultralytic company. It is based on YOLOv4 and features three sections (backbone, neck, and head) but lacks SPP in the neck, while implementing an additional Logit loss function.

Although the YOLOv5 detection method is well established for detection application, its effect on 1D and 2D attention models in YOLOv5 for WBCs type detection has not been extensively studied. Thus, this research studied YOLOv5 detection models for four types of WBCs by training YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x with 2800 train images from the public dataset Kaggle. The fusion of attention model SE and CBAM into YOLOv5s was investigated and the results were analyzed. Subsequently, the performance of the developed models was tested on the BCCD dataset.

1.1 Automated WBCs Detection and Classification

The automated detection method of WBCs in the microscopic blood smear images has evolved throughout the years. Computer aided automated detection methods have become the trend in WBCs related disease such as leukaemia detection method in studies [12-14]. In a study by Tahiri *et al.,* [15] used the combination of discrete moment quaternion, artificial intelligence, and machine learning approaches for classification of fundus images. These methods include image processing, computer vision and artificial intelligent techniques. Several WBCs deep learning detection method studies had presented very promising results. In a study by Novoselnik *et al.,* [16], the convolutional neural network and modified LeNet-5 method is used for detecting and classifying the five types of

WBCs. The overall accuracy result obtained is 81.11%. In comparison, a study on a one-stage WBC detection method that was based on the YOLOv2 model [17]. It is improved by adding the multi-scale features and Fourier ptychographic microscopy (FPM). It showed a result of 1.00 precision and recall value. Later, CNN algorithm with combination of minimum redundancy maximum relevance (MRMR) feature selection and extreme learning machine (ELM) classifier was used in a study and achieved 96.03% of accuracy for WBC detection on BCCD dataset [18].

A CNN and region proposal approach for WBC detection and counting have been studied and obtained the accuracy of 97% [19]. Other than that, a K-means cluster and CNN-VGG16 classifier method was developed for five types of WBC detection and classification. The model was trained and tested on clinically healthy canine blood smear images. The overall accuracy of the proposed method is 95.89% [20]. In the same year, a higher accuracy regional convolutional neural network (RCNN) together with the ResNet50 method was developed for detection of the WBC types [21]. A CNN based detection also has been implemented in detecting acute lymphoblastic leukaemia cells [22]. The CNN model consists of ten layers with convolutional layers, maximum pooling layers, leaky ReLu layers, fully connected layers, flatten layer, dropout layer, SoftMax layer, and lastly the classification output layer. Dataset used for training was from the Acute Lymphoblastic Leukaemia Image Database (ALL-IDB). Machine learning based for detecting abnormality in WBC was developed using K-means clustering, K-nearest neighbour (KNN) and support vector (SVM) [23]. The images were pre-processed before fed into the designed machine learning model. A Faster-RCNN model was utilised in a study for immature blood cell detection [24]. It is divided into five types of immature blood cells, and the blast cell achieved the highest accuracy of 93.884%. A paper had studied the detection and classification technique using multilayer AlexNet feature selection (MLANet-FS) to identify the WBC type [25]. The combination with ELM (Relief) classifier and attained the accuracy of 99.12%.

A paper studying a one-stage WBC detection using YOLOv3 gained the accuracy of 99.2% [26]. It is able to detect the four types of WBCs which are the eosinophil, lymphocyte, monocyte, and neutrophil with precision values of 1.00, 1.00, 1.00, and 0.97 respectively. A paper also did a study on YOLOv3 based WBCs detection with LISC dataset [27]. The Darknet feature extraction in the original YOLOv3 model was replaced with a much simpler CNN model, AlexNet. This replacement achieved accuracy of 98%. Another paper that used YOLOv3 together with the improved squeeze-excitation (SE) attention module for blood cell detection that include the WBC, RBC and platelets [28]. This application attained 85.7% overall accuracy and 95.6% for WBCs, 92.7% for RBCs and 89.6%. As the YOLO detection model keeps developing new versions, a study on the YOLOv4 for acute lymphoblastic leukaemia diagnosis has been made [29]. The study was done on the datasets from ALL-IDB1, and C_NMC_2019 and the accuracy is 96.06% and 98.7%. The later version of YOLO series is the YOLOv5 and a detection study on blood cell type (WBC, RBC, and platelets) have been made [30]. It is done on the BCCD dataset with 364 images. The results show 93% of overall accuracy. In addition, training with a bigger dataset size had improved the detection accuracy. This is shown in a study using the same YOLOv5 model but with 2800 WBC dataset from Kaggle.com [31]. It gained an overall accuracy of 99.42%. Then, a modification from the YOLOF model has been made by a paper of Xu *et al.,* [32]. It is shown to be a light-weight model which implements EfficientNet in the backbone of the YOLOF model. It is able to detect WBC, RBC and platelets with 91.9% mAP value.

A latest CNN based detection that integrates Faster R-CNN with EfficientDet D3 [33]. The accuracy result is high, with 99.4% detection accuracy on WBCs, RBC, and platelets dataset. However, the author claimed that the training time and inference time were slower in comparison with the recent one-stage detector such as the YOLO series. An alternative approach by a researcher developed a deep feature convolutional neural network (DFCNN) together with AlexNet, GoogleNet, and ResNet50 for feature extraction [34]. The hybrid mayfly algorithm (HMA) and particle swarm

optimization (PSO) were applied as feature selection for the overall model. The selected features are then classified by Recurrent Neural Network- Long Short-Term Memory (RNN-LSTM) into four WBCs class types. The overall accuracy of the system is 97%.

Most of the recent WBC detection methods developed by researchers have obtained an accuracy of above 90%. It is also able to train and detect up to five numbers of classes of WBCs type and leukaemia cells. Other than that, the dataset commonly used for WBC detection and counting is from BCCD. Whilst for WBC type detection, dataset from Kaggle.com were used. Leukaemia cells datasets were also obtained from the public dataset which are from the ALL-IDB and TCIA Online Database.

1.2 Attention Modules

Attention model in artificial intelligence was originally adapted based on human cognitive ability to focus on only important details in detecting processes [35,36]. For example, in Figure 1 a class photo with three rows of students and a teacher. If we are asked to locate the teacher of the class group, our brain will automatically first search the centre and most front rows and ignore other parts of the photo information. The brain will detect the outfit differently in colour and style to verify it. The same method applies to our brain. If we are looking for a boy student named Ali who has short hair and is wearing spectacle, the brain will only search for short hair and the one that is wearing spectacle. That is a general concept of attention. This concept was then implemented in the artificial intelligent application in visual object detection.

Specifically in deep learning object detection applications, this method was claimed to be more robust. CNN base attention module can be divided into three types. The channel attention, spatial attention, and the mixed attention. The purpose of the channel attention is to link between each channel by modelling and acquire the significance in each feature channel by network learning. The strength of significance is assigned by weight value for each channel and the non-significant feature is suppressed. The SE attention model is one of the channel type attention modules [37]. It consists of a squeeze and excitation part. The squeeze part will reduce the dimension then through the dimension expansion. This then will compress the value of the parameter and improve the efficiency of the calculation. The module uses two multilayer perceptron for correlation learning between channels. Whereas the spatial attention module is when the spatial feature information in a channel is assigned by weight value.

A simple and effective attention module was designed for a feed-forward convolutional neural network called CBAM [38]. The module performs mapping by channel and also by spatial in a sequence. Then, for the feature refinement, the attention maps are multiplied with the input feature map. This lightweight and general model is able to be integrated into any CNN structure with minimal overheads and trainable together with the base CNNs. The implementation of CBAM on detection dataset (ImageNet-IK, MS-COCO and VOC 2007) with various CNNs base models generate a consistent performance for classification and detection tasks.

A different attention model is the style-based recalibration module (SRM) [39]. It recalibrates the intermediate feature maps by style. Style pooling consists of average pooling and global standard deviation pooling operated in parallel and used to extract style features from each channel. The recalibration weight is estimated by channel-independent style integration which is the process of channel wise fully connected layer, batch normalisation and sigmoid function. The use of SRM will enhance the CNN representation ability with negligible overhead.

2. Methodology

2.1 Datasets Acquisition and Preparation

The WBCs public dataset was used for this experiment. It was obtained from kaggle.com and divided into 2800 training images, 786 for validation, and 405 images for model testing which from the spilt ratio of 70:20:10. The images include the four types of WBC (eosinophil, lymphocytes, monocytes, and neutrophil). Figure 1 is the sample images from the dataset. The annotation and managing the dataset was done in Roboflow.com. To prepare the dataset for the detection process, an annotation process was conducted and the Roboflow platform was used to facilitate this task. As for the distribution of image types for each class, there were 980, 992, 987, 983 for eosinophil, lymphocyte, monocyte, and neutrophil respectively. Basophil type was not considered in this experiment because it is rarely found in blood smear images due to its 1% composition of the white blood cells.

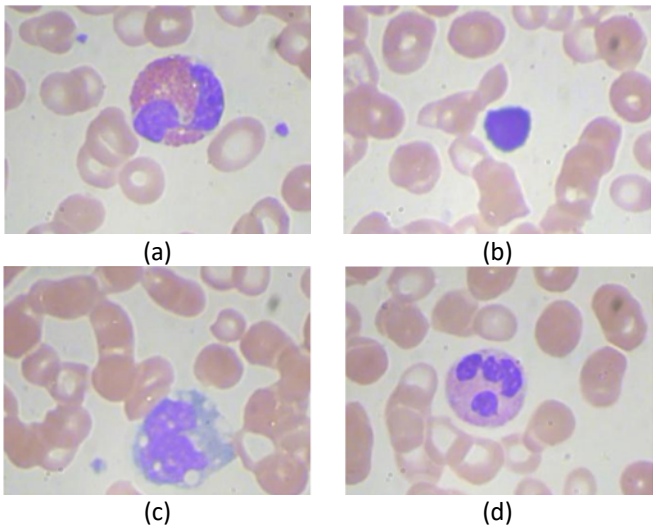

Fig. 1. Microscopic image of (a) Eosinophil, (b) Lymphocyte, (c) Monocyte, and (d) Neutrophil

2.2 Training Environment and Tools

This research performed the experiment using a laptop with specification of the Intel® Core™ i5-5200U CPU @ 2.20GHz processor. Kaggle.com/code and Google Colab with GPU are used for programming and running the experiment. Pytorch is the main framework used for the YOLOv5, SE and CBAM coding. The reference for YOLOv5 was obtained through the Ultralytic github web page.

2.3 YOLOv5s

This research developed a WBC type detection using improved YOLOv5s with SE and CBAM attention modules. YOLOv5s is the lightest version from the YOLOv5 series. The attention modules were proposed to increase the accuracy of the model while maintaining its lightweight characteristic. YOLOv5s model structure consists of backbone, neck, and the head which illustrated in Figure 2.

The model starts with the backbone which the training images were fed first. In the backbone, there are steps of convolutional layers, bottleneckCSP (labelled as C3 in Figure 2 and Figure 3) layers and the SPPF layer. The bottleneckCSP is a cross stage partial network that is based on the DenseNet

[40]. The input will be divided into two. It was designed to cater the gradient information duplication problem in the network. In addition, it is able to reduce the model complexity while having the same accuracy. Then, the SPPF is a faster version of the Spatial Pyramid Pooling (SPP) [41].

The next segment of the YOLOv5 model is the Neck part. This is where the feature extraction procedure takes place. Here, there are sets of convolutional layers, concatenation, upsampling operation and the bottleneckCSP. A bottom-up augmentation path, known as the Path Aggregation Network (PAN) was adapted in this section. It is to make the propagation of the low-layer information easier [42]. The adaptive pooling feature allowed information from all levels to be accessed by each proposal for prediction.

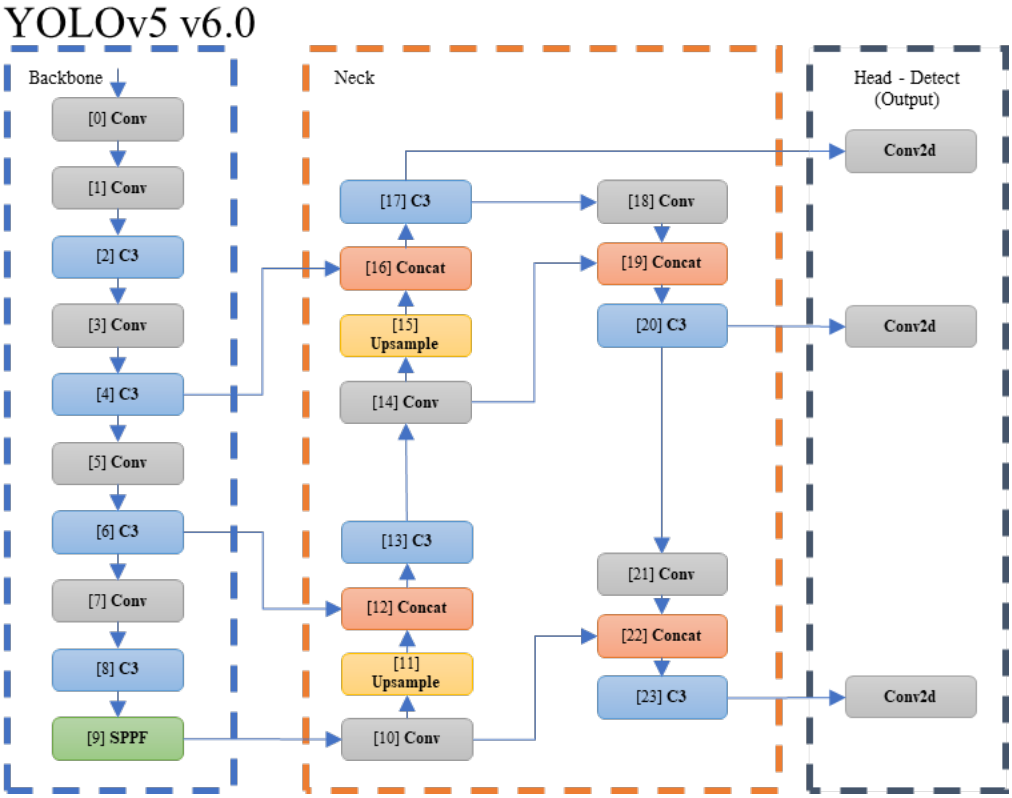

Fig. 2. YOLOv5 v6.0 model structure

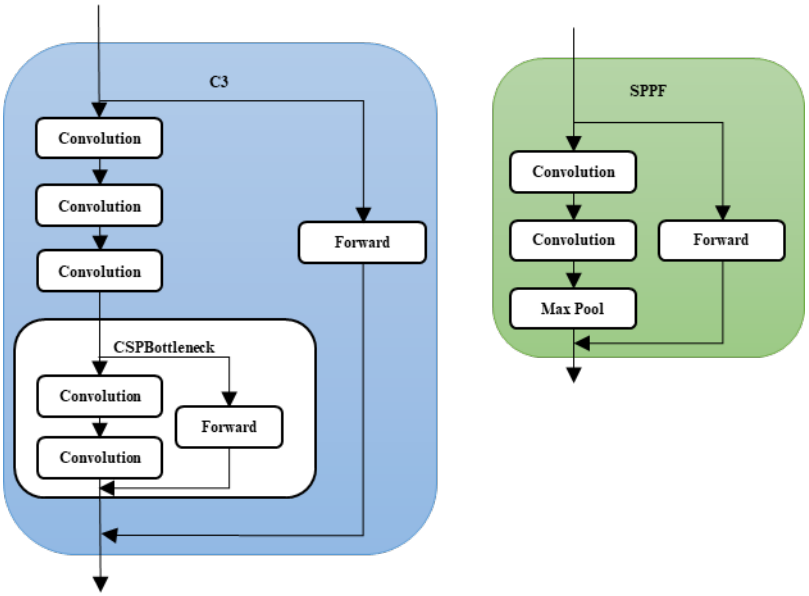

Fig. 3. Sub structure of C3 and SPPF

The concatenation operation in the neck section connects the feature map of the PAN. Then, the outputs are fused together by the fully connected layers. The bounding box and class prediction will then be performed in the Head section where the bounding box coordinates, class labels, and confidence score will be generated.

The difference in each YOLOv5 series is the number of layers (model depth). The higher the number of layers, the deeper the model. The YOLOv5s has the smallest number of layers, 283. On the other hand, the YOLOv5x has the largest number of layers of 607 with 87257832 parameters. The details are tabulated in Table 1.

Table 1
Parameter Size for Each YOLOv5

YOLOv5 models	Layers	Parameters
YOLOv5s (small)	283	7068936
YOLOv5m (medium)	391	21064488
YOLOv5l (large)	499	46642120
YOLOv5x (extra-large)	607	87257832

The research procedure starts with dataset labelling and organization. The labelled train images are then fed into the backbone, the neck, and the head where the output is generated. After the training process is complete, the obtained weight is used for testing and inferences. The process is illustrated in Figure 4. Figure 5 illustrates the feature extraction and identification process of WBC in the YOLOv5. The input image was fed into the backbone for extracting the object features in image using BottleneckCSP that includes CSP and SPP which splices as s x s grid and stacks units. In the Neck part, down-sampling was performed for feature extraction and up-sampling process to restore the feature map size as in the FPN. This process produced multi scale predictions and fed to head part of the YOLOv5 where the probability mapping, bounding box prediction, and confidence score were calculated. Non maximum suppression algorithm was used for prediction box filtering and generated the final detection result.

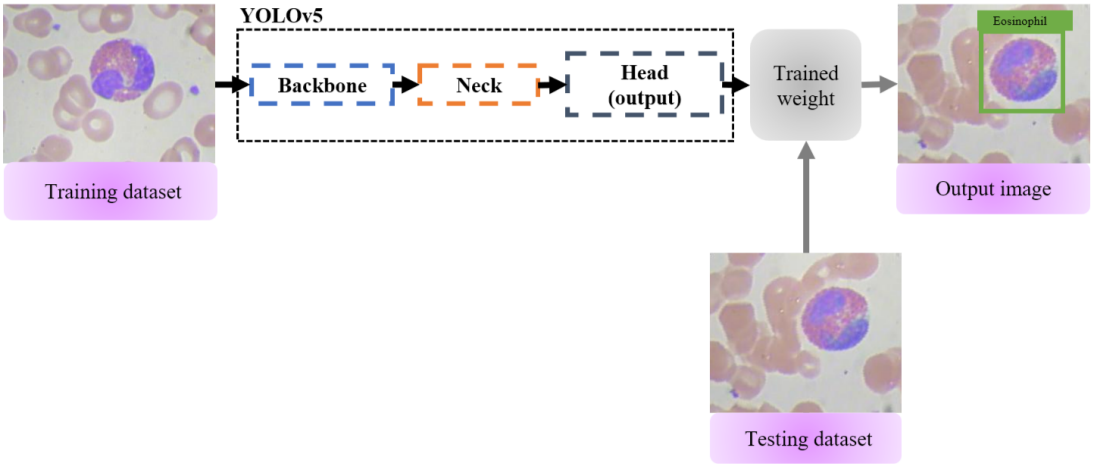

Fig. 4. Application of YOLOv5 model for WBCs detection

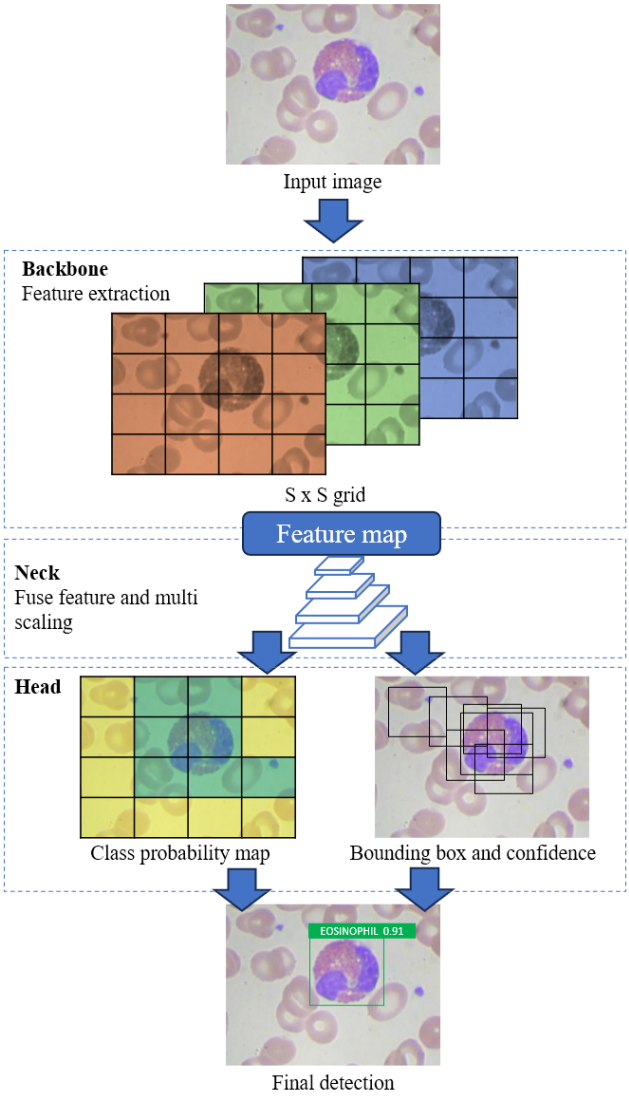

Fig. 5. Feature extraction and identification of WBC

This paper proposed the WBCs type detection method using YOLOv5s which include the attention models. Based on the previous experiments, the YOLOv5m achieved the most accurate mAP value

compared to other YOLOv5 series. This makes a benchmark result for improving the lighter model, which is the YOLOv5s. In order to do so, the attention mechanism was implemented in the model. The attention models that have been investigated are the squeeze-excitation (SE) and the convolutional block attention module (CBAM).

2.4 Squeeze-Excitation (SE) Model

Squeeze-excitation was designed for feature recalibration. The interrelationship between feature channels was modelled by weight of different feature channels. The first part of the model is the squeeze operation. It aggregated the feature maps across its spatial dimensions and produced a channel descriptor. The feature dimension of H×W×C is compressed into 1×1×C. Excitation operation acts like a simple self-gating mechanism. It then produces the weights for each channel which then applied to the feature maps U and generates the output for the SE module. Figure 6 shows the model structure of the SE attention model.

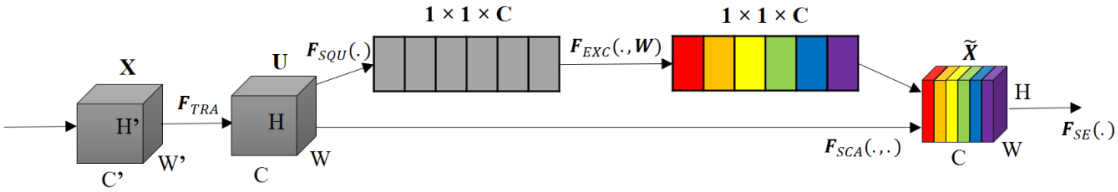

Fig. 6. Squeeze-Excitation attention model structure

The squeeze function equation is as in Eq. (1).

$$F_{SQU}(u) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{1}$$

The excitation function is derived as in Eq. (2), Eq. (3), Eq. (4) and Eq. (5). W_fc1 and W_fc2 are the value for the first and second fully connected layer weight. The ReLU activation represent in Eq. (3) and δ( ) is the sigmoid activation function.

$$F_{fc1} = W_{fc1} * F_{SQU}(u_c) \tag{2}$$

$$F_{ReLU} = \varphi(F_{fc1}) \tag{3}$$

$$F_{fc1} = W_{fc2} * F_{ReLU} \tag{4}$$

$$F_{EXC} = \left(F_{SQU}(u_c)\right) = \delta(F_{fc2}) \tag{5}$$

The SE function can be obtained by matrix multiplication of input feature map with the output of from the excitation module as in Eq. (6).

$$F_{SE} = F_{EXC}\left(F_{SQU}(u_c)\right) \otimes u_c \tag{6}$$

The total number introduced by weight parameters of these FC layers is given by:

$$\frac{2}{r}\sum_{s=1}^{S} N_s . C_s^2 \quad (7)$$

Where the abbreviation, r = reduction ratio, S = number of stages, Cs = output channel dimension, Ns = number of repeated blocks in stage s.

2.5 Convolutional Block Attention Module (CBAM)

The CBAM considers both channel and spatial feature information. The first part of the module is the channel feature information extraction. It generates features from the average pooling, F_{avg}^{C} and maximum pooling, F_{max}^{C} methods. Both the features extracted then combine to a shared network, $M_C(F)$ which also consist of the multi-layer perceptron, MLP. Then, the output features were combined using element summation. This operation is shown in Eq. (8), where W_0 and W_1 are the MLP weight. σ is the sigmoid function and ReLU is the activation function.

$$M_C(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) = \sigma\left(W_1\left(W_0\left(F_{avg}^{C}\right)\right) + W_1\left(W_0\left(F_{max}^{C}\right)\right)\right) \quad (8)$$

Next is the spatial part of the module which determines whether to emphasise or to suppress information. A spatial attention map is generated by Eq. (9), where the average pooling and maximum pooling operation will generate 2D maps. Those pooling outputs then concatenated and convolved to produce 2D spatial attention map. Its equation is as Eq. (9), where σ is the sigmoid function and $f^{7\times7}$ is the convolutional operation with 7 by 7 filter size.

$$M_S(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right) = \sigma\left(f^{7\times7}\left(\left[F_{avg}^{S}; F_{max}^{S}\right]\right)\right) \quad (9)$$

The channel attention and spatial attention then run sequentially to construct refined output features. Figure 7 illustrates the process of the CBAM.

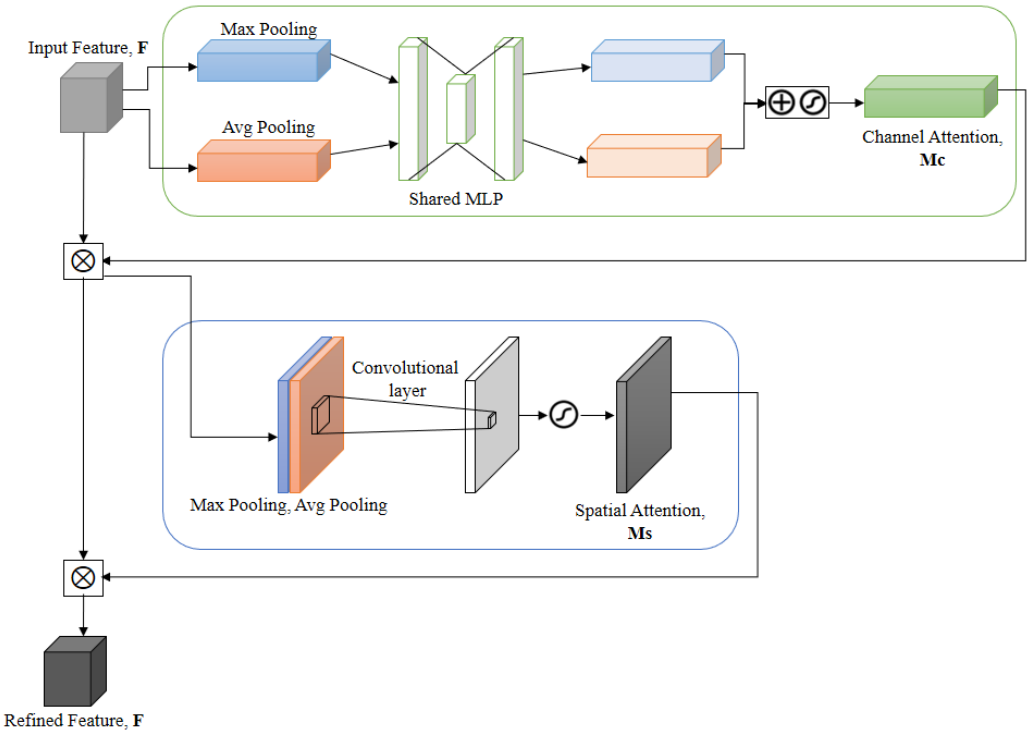

Fig. 7. CBAM model structure

2.6 Measurements and Evaluation

The evaluation of the experiment was done by calculating the accuracy, precision and sensitivity value. The equations are as in Eq. (10), Eq. (11) and Eq. (12).

$$Accuracy\ (Acc.) = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$Precision\ (Pre.) = \frac{TP}{TP+FP} \tag{11}$$

$$Sensitivity\ or\ Recall\ (Rec.) = \frac{TP}{TP+FN} \tag{12}$$

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative. For overall performance, the mean average precision (mAP) value is calculated as in Eq. (13) and Eq. (14). The AP value is calculated by differentiating the function Pre.(Rec.) with respect to the Rec. the limit is set between 0 to 1 and k is the total number of classes.

$$AP = \int_0^1 Pre.(Rec.)\, dRec. \tag{13}$$

$$Mean\ Average\ Precision\ (mAP) = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{14}$$

The output images display a detection box labelled with the WBCs type name and the confidence value which represents the level of confidence that the bounding box contains an object. The confidence value is calculated using Eq. (15).

$$Confidence = p_r(Object)\ \times IoU_{pred}^{truth},\ \ p_r(Object) \in \{0,1\} \tag{15}$$

Where $p_r(Object)$ is the probability of object, IoU_{pred}^{truth} is the intersection over union between the predicted box and ground truth box. The mAP0.5 indicates that the IoU threshold is 0.5, while mAP 0.5:0.95 refers to when the threshold is set from 0.5 to 0.95.

The model classification performance evaluated by calculating its Cohen kappa statistic, κ and Matthew's Correlation Coefficient, (MCC). Eq (16) shows the calculation for kappa statistic while Eq. (17) shows the MCC's equation.

$$\kappa = \frac{2*(TP*TN)-(FN*FP)}{(TP+FP)*(FP+TN)+(TP+TN)*(FN+TN)} \tag{16}$$

$$MCC = \frac{TN*TP-FN*FP}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{17}$$

2.7 Hyperparameter Setup

The hyperparameter setup are tabulated as in Table 2 and most of the parameters were in default YOLOv5 mode. The final learning rates were set lower, and the reduction ratio were introduced in fusion method.

Table 2
Hyperparameters setup

HYPERPARAMETERS	YOLOV5S	YOLOV5SSEC2	YOLOV5SCBAM2BCK	YOLOV5SSE1CBAM2
HYP.BOX	0.05	0.05	0.05	0.05
HYP.CLS	0.5	0.5	0.5	0.5
HYP.LR0	0.01	0.01	0.01	0.01
HYP.LRF	0.2	0.01	0.01	0.01
HYP.OBJ	1	1	1	1
HYP.HSV_H	0.015	0.015	0.015	0.015
HYP.HSV_S	0.7	0.7	0.7	0.7
HYP.HSV_V	0.4	0.4	0.4	0.4
HYP.IOU_T	0.2	0.2	0.2	0.2
HYP.SCALE	0.5	0.5	0.5	0.5
HYP.ANCHOR_T	4	4	4	4
HYP.FL_GAMMA	0	0	0	0
HYP.MOMENTUM	0.937	0.937	0.937	0.937
HYP.TRANSLATE	0.1	0.1	0.1	0.1
HYP.WEIGHT_DECAY	0.0005	0.0005	0.0005	0.0005
IMGSZ	416	416	416	416
EPOCHS	100	100	100	100
DATA_DICT.NC	4	4	4	4
BATCH_SIZE	16	16	16	16
BBOX_INTERVAL	-1	-1	-1	-1
REDUCTION RATIO	N/A	8/16/32	16	32

3. Result

3.1 YOLOv5s-SE

This section documented the implementation of the SE module in the YOLOv5s model. The experiments were divided into two parts. Firstly, is to investigate how the value of the reduction ratio affects the WBC type detection, and secondly the effect of the SE positioning in the YOLOv5s on WBC detection. The mean average precision value is used to compare the result obtained. The SE module was placed in the backbone of the YOLOv5s. Figure 8 illustrates the YOLOv5s-SE fusion.

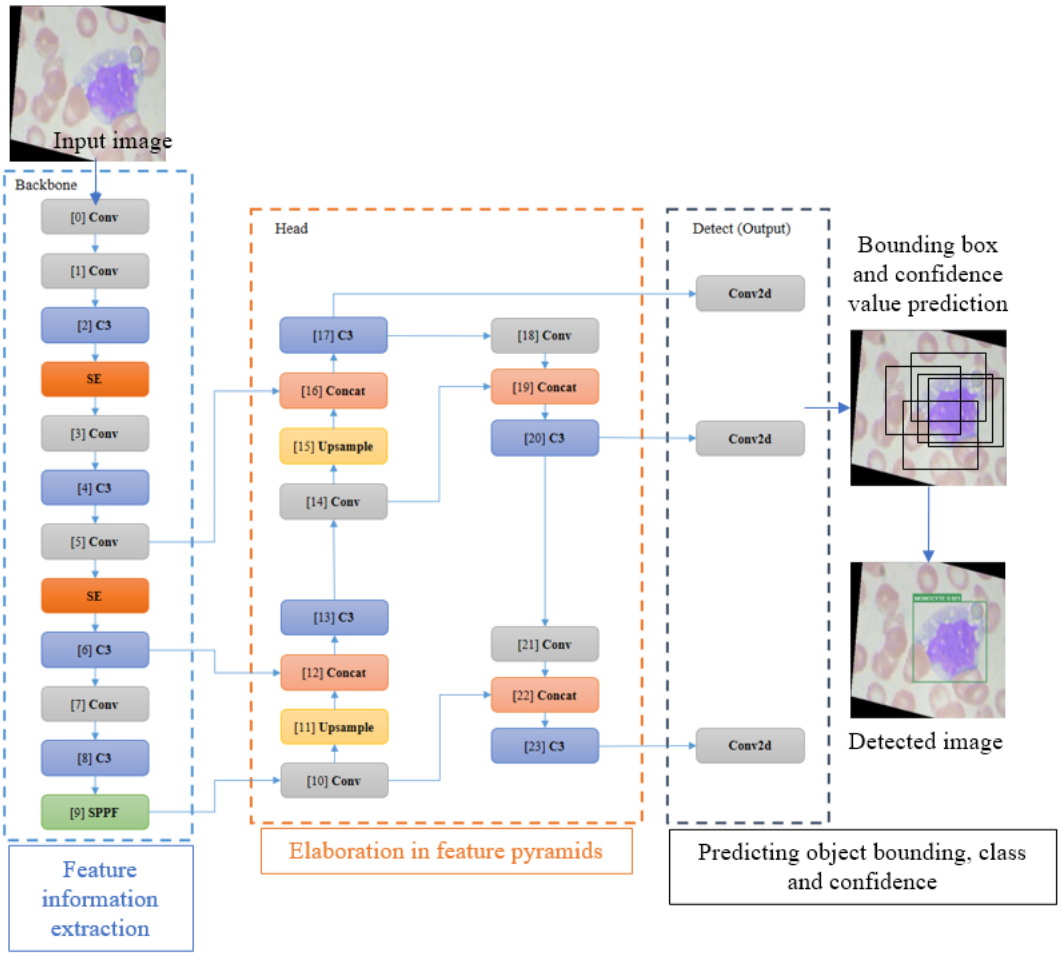

Fig. 8. YOLOv5s with SE model

3.1.1 SE model in YOLOv5s with different reduction ratio

The SE model is incorporated inside the YOLOv5 structure. It is initially placed in the fourth layer of the backbone. Then, the reduction ratios varied by 8, 16 and 32. The reduction ratio, r in the SE model was implied in the dimensional-reduction layer.

The training result was tabulated in Table 3, which includes the results of mAP 0.5 value, mAP 0.5:0.95 value, training time, detection speed and the model parameter. It can be concluded that the reduction ratio of 32 obtained the highest mAP value. However, it took longer to train and detect.

Table 3
Training result for SE model in YOLOv5s with different reduction ratio value

Reduction ratio, r	mAP 0.5 (%)	mAP 0.5:0.95 (%)	Training time (hour)	Detection Speed	Parameter
Without SE	99.29	81.7	0.851	8.7ms, 0.9ms/image	7068936
8	99.3	82.3	0.832	8.3ms, 0.8ms/image	7209105
16	99.3	82.3	0.840	8.7ms, 0.9ms/image	7209105
32	99.4	82.6	0.901	9.0ms, 1.0ms/image	7209105

3.1.2 SE model position in YOLOv5s

The second SE model was included in the YOLOv5 structure. Placement of SE model then tested by placing it in three different configurations as in Table 4. Configuration 2 has the highest mAP0.5 value of 99.5 and shortest training time 0.895hour, while configuration 1 has the highest mAP0.5:0.95 value of 82.5. The detection speed for both configurations remain the same, which is 0.9ms per image.

Table 4
Training result for SE model position in YOLOv5s

Configuration	Position	mAP 0.5 (%)	mAP 0.5:0.95 (%)	Training time (hour)	Detection speed	Parameter
0	Without SE	99.29	81.7	0.851	8.7ms, 0.9ms/image	7068936
1	SE (after conv5)	99.4	82.5	0.906	9.1ms, 0.9ms/image	7209105
2	SE-PRE (after conv2)	99.5	82.3	0.895	9.1ms, 0.9ms/image	7208209
3	SE, SE-PRE (conv2, conv5)	99.3	81.7	0.905	9.1ms, 1.0ms/image	7250577

3.2 YOLOv5s-CBAM

The incorporation of the CBAM in YOLOv5 was then tested with different position configurations in the backbone and the neck of the YOLOv5 structure. The experiment was designed to test four discrete position configurations. The first configuration was by placing one CBAM in the backbone. Second was placing one CBAM in the neck. Two CBAM placed in the neck for the third configuration. Lastly, the combination of all positions which one CBAM in the backbone and two CBAM in the neck.

The outcome of the experiment was tabulated in Table 5. It shows that configuration 1 obtained the highest mAP0.5 and mAP0.5:0.95 values which are 99.5% and 82.6% respectively. On top of that, it also achieved the fastest detection speed of 0.8ms per image. Figure 9 illustrates the YOLOv5s-CBAM fusion.

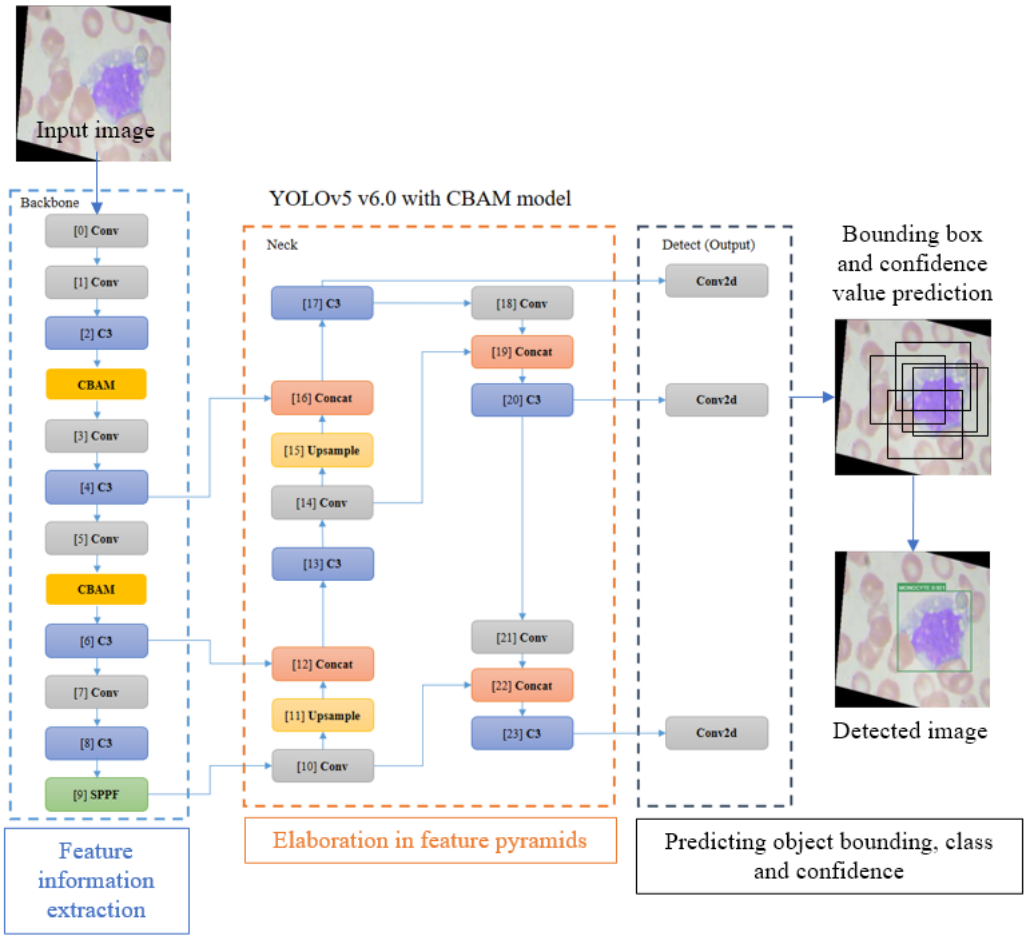

Fig. 9. YOLOv5s with CBAM model

Table 5
Training result for CBAM model position in YOLOv5s

Configuration	Position	mAP 0.5 (%)	mAP 0.5:0.95 (%)	Training time (hour)	Detection speed	Parameter
0	Without CBAM	99.29	81.7	0.851	8.7ms, 0.9ms/image	7068936
1	CBAM in Backbone (2 models)	99.5	82.3	0.938	8.9ms, 0.8ms / image	7198870
2	CBAM in Neck (1, after first concat)	99.4	82.5	0.855	9.0ms, 0.9ms / image	7135093
3	CBAM in Neck (2, after first `& second concat)	99.3	82.6	0.894	10.2ms, 0.9ms / image	7129833
4	CBAM in Backbone and Neck	99.4	80.9	0.949	11.2ms, 0.9ms / image	7323665

3.3 YOLOv5s-SE-CBAM

The next experiment is the combination of SE and CBAM models fusion into the YOLOv5s model. Figure 10 illustrated the employment of the SE and CBAM model in the backbone and the neck

structure. CBAM models are placed each in the backbone and neck, while the SE model was first placed into the backbone only, then it was placed in both backbone and neck part of the YOLOv5 structure. The results are shown in Table 6. The SE1-CBAM2 has the highest mAP0.5 which is 99.4%, but the lowest mAP0.5:0.95 of 81.0%.

Table 6
Training result for SE-CBAM model position in YOLOv5s

Model	mAP 0.5 (%)	mAP 0.5:0.95 (%)	Training time (hour)	Detection speed	Parameter
Without SE-CBAM	99.29	81.7	0.851	8.7ms, 0.9ms/image	7068936
SE1-CBAM2	99.4	81.0	0.978	10.9ms, 0.9ms / image	7323021
SE2-CBAM2	99.3	81.8	0.948	10.9ms, 0.9ms / image	7379245

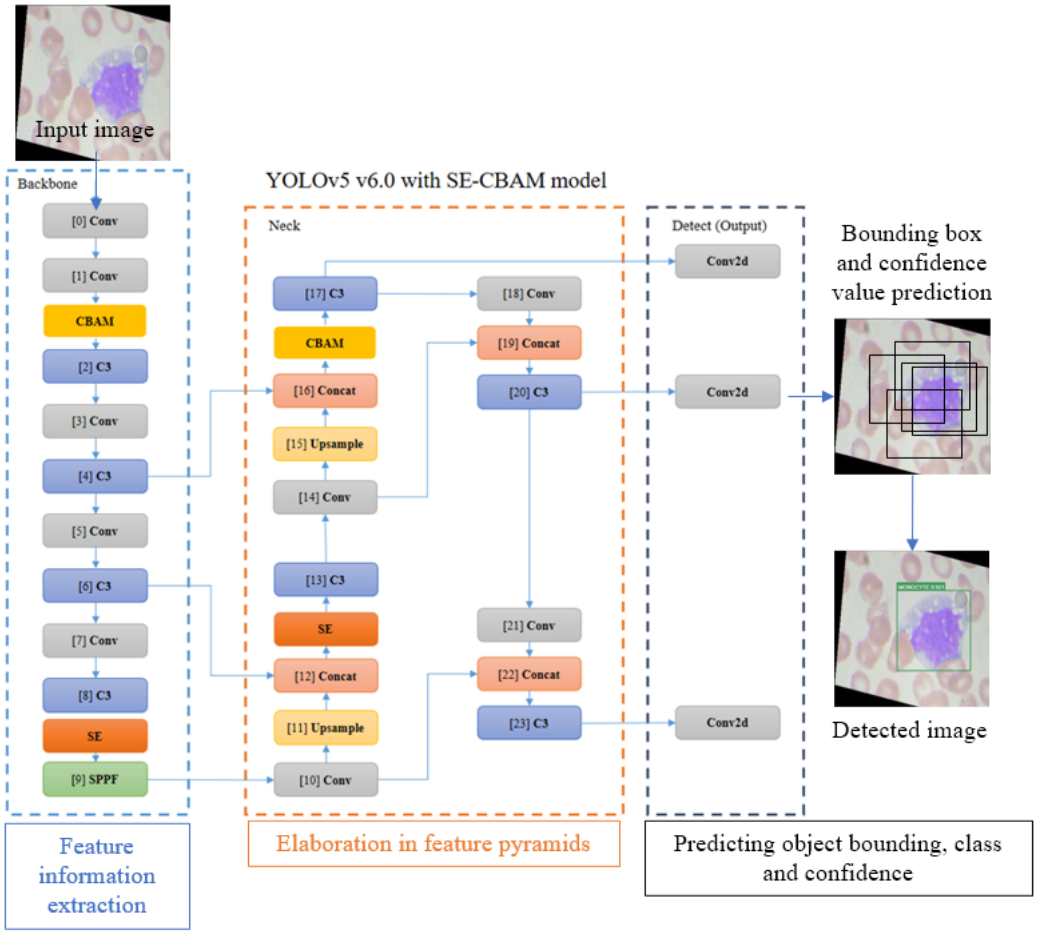

Fig. 10. YOLOv5s with SE and CBAM models

Table 7 compared the overall results for YOLOv5s, YOLOv5sSE, YOLOv5sCBAM and YOLOv5sSECBAM in terms of the accuracy, precision, recall, kappa, and MCC value.

Table 7
Overall results for the developed models

	YOLOV5V	YOLOV5SSE	YOLOV5SCBAM	YOLOV5SSECBAM
ACCURACY	0.9929	0.9946	0.9937	0.9939
PRECISION	0.9868	0.9937	0.989	0.9958
RECALL	0.9757	0.9918	0.9897	0.9935
κ	0.4611	0.2715	0.2623	0.0644
MCC	0.4673	0.2728	0.2623	0.0658

3.4 Comparison of WBC Detection Methods

In comparison, the mAP value of the proposed models ranges from 99.1% to 99.5% for each type of the WBC and overall value between 99.4% and 99.5% as shown in Table 8. It is higher than other CNN based detection methods. Even though the result is on par with the YOLOv5m, the proposed methods have lower parameters which make them lightweight and reliable detection model for WBCs type.

Table 8
Comparison of WBCs type detection methods

		Accuracy (%)					Parameter
Year / Study	Method	*Eosinophil*	*Lymphocyte*	*Monocyte*	*Neutrophil*	*Overall*	
2020 / [17]	K-Means cluster & CNN-VGG16	93.36	98.70	95.50	99.59	95.89	-
2020 / [18]	RCNN + ResNet50	96.16	99.52	98.40	95.04	97.52	-
2020 / [24]	YOLOv3, AlexNet	100.00	100.00	90.00	100.00	98.00	-
2021 / [22]	MLANet-FS-ELM (ReliefF)	97.42	100.00	99.68	99.35	99.12	-
2022 / [28]	YOLOv5m	99.3	99.6	99.6	99.2	99.4	21064488
2022 / [28]	YOLOv5s	99.0	99.6	99.5	99.0	99.3	7068936
Proposed	YOLOv5s-SEc2	99.4	99.5	99.5	99.4	99.5	7198705
Proposed	YOLOv5s-CBAM	99.5	99.5	99.5	99.4	99.5	7198870
Proposed	YOLOv5s-SE1-CBAM2	99.4	99.5	99.5	99.1	99.4	7323021

4. Conclusion

This research performed the training for the YOLOv5s structure with the integration of SE and CBAM attention models. 2800 microscopic images of WBCs type, eosinophil, monocyte, lymphocyte, and neutrophil were used for training. Several possible combinations were tested, and the experiment results showed that the application of the attention module in YOLOv5s had improved its accuracy while maintaining its lightweight. The proposed configuration YOLOv5s-SE, YOLOv5s-CBAM, and YOLOv5s-SE-CBAM achieved promising results where the obtained mAP values are 99.5%, 99.5% and 99.4% respectively. The accuracy had increased while maintaining lightweight detection models which the parameters value is 65.8% lower than the YOLOv5m.

Overall, the contribution of this study included:

(i) This research has contributed to lightweight WBC type detection by fusing attention models into the YOLOv5s model, making it suitable for real-time applications and capable of performing efficiently on low-resources devices with lower GPU power consumption. This results in a cost-effective and low-maintenance solution.

(ii) The developed fusion model, YOLOv5sCBAM2bck, achieves a higher detection rate of 0.8ms per image, which aids pathologists in quickly detecting WBCs.

This research limitations are:

(i) The detection models were trained on 2800 images, validated on 786 images and tested on 405 images from Kaggle.com public dataset. The final comparison was conducted using 44 WBCs images from the BCCD dataset.
(ii) This research focused on detecting four types of WBCs: Eosinophil, Lymphocyte, Monocyte and Neutrophil.
(iii) The experiments were performed using Google Colab for coding, training and testing.
(iv) The coding was done in the Python programming language by using Keras, Tensorflow and Pytorch frameworks.
(v) The detection model structure was based on the YOLOv5 detection model, including YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x.
(vi) This research focused on fusing SE and CBAM attention model into YOLOv5s.
(vii) This research was conducted on a Dell Laptop with an Intel® Core™ i5-5200U CPU @ 2.20GHz processor.

The WBCs detection model, which uses a fusion of YOLOv5s and an attention model, holds promise for future research. It could be further explored and tested with different datasets such as Raebi and LISC to evaluate its performance across various datasets. Additionally, there is potential to train the proposed model with abnormal WBCs microscopic image datasets such as the ALL-IDB dataset, which contains images of ALL cells.

Acknowledgement
This research was funded by a grant from Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2019/TK04/UTHM/02/6).

References

[1] Nicholson, Lindsay B. "The immune system." *Essays in Biochemistry* 60, no. 3 (2016): 275-301. https://doi.org/10.1042/EBC20160017
[2] Martin, Bridget. *Harmful interaction between the living and the dead in Greek tragedy*. Liverpool University Press, 2020. https://doi.org/10.3828/liverpool/9781789621501.001.0001
[3] Bain, B. J. "The nature of leukaemia, cytology, cytochemistry and the FAB classification of acute leukaemia." *Leukaemia Diagnosis*, 4th edition, Wiley-Blackwell, Chichester, UK (2010): 1-63. https://doi.org/10.1002/9781444318470.ch1
[4] Parveen, R., M. Nabi, F. A. Memon, S. Zaman, and M. Ali. "A review and survey of artificial neural network in medical science." *Journal of Advanced Research in Computing and Applications* 3, no. 1 (2016): 7-16.
[5] Anaya-Isaza, Andrés, Leonel Mera-Jiménez, and Martha Zequera-Diaz. "An overview of deep learning in medical imaging." *Informatics in Medicine Unlocked* 26 (2021): 100723. https://doi.org/10.1016/j.imu.2021.100723
[6] Li Wen, Khaw and Shah Abdullah, Shahrum. "MRI Brain Image Classification Using Convolutional Neural Networks and Transfer Learning." *Journal of Advanced Research in Computing and Applications*, 31, Issue 1(2023): 20-26.
[7] Sanchez, S. A., H. J. Romero, and A. D. Morales. "A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework." In *IOP Conference Series: Materials Science and Engineering*, vol. 844, no. 1, p. 012024. IOP Publishing, 2020. https://doi.org/10.1088/1757-899X/844/1/012024
[8] Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788. 2016. https://doi.org/10.1109/CVPR.2016.91
[9] Redmon, Joseph, and Ali Farhadi. "YOLO9000: better, faster, stronger." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263-7271. 2017. https://doi.org/10.1109/CVPR.2017.690
[10] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).

[11] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
[12] Zhang, Xunxun, and Xu Zhu. "Vehicle detection in the aerial infrared images via an improved yolov3 network." In *2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP)*, pp. 372-376. IEEE, 2019. https://doi.org/10.1109/SIPROCESS.2019.8868430
[13] Bagasjvara, R. G., Ika Candradewi, Sri Hartati, and Agus Harjoko. "Automated detection and classification techniques of Acute leukemia using image processing: A review." In *2016 2nd International Conference on Science and Technology-Computer (ICST)*, pp. 35-43. IEEE, 2016. https://doi.org/10.1109/ICSTC.2016.7877344
[14] Alsalem, M. A., A. A. Zaidan, B. B. Zaidan, M. Hashim, Osamah Shihab Albahri, Ahmed Shihab Albahri, Ali Hadi, and K. I. Mohammed. "Systematic review of an automated multiclass detection and classification system for acute Leukaemia in terms of evaluation and benchmarking, open challenges, issues and methodological aspects." *Journal of Medical Systems* 42 (2018): 1-36. https://doi.org/10.1007/s10916-018-1064-9
[15] Tahiri, Mohamed Amine, Hicham Amakdouf, Mostafa El Mallahi, and Hassan Qjidaa. "Optimized quaternion radial Hahn Moments application to deep learning for the classification of diabetic retinopathy." *Multimedia Tools and Applications* 82, no. 30 (2023): 46217-46240. https://doi.org/10.1007/s11042-023-15582-9
[16] Novoselnik, Filip, Ratko Grbić, Irena Galić, and Filip Dorić. "Automatic white blood cell detection and identification using convolutional neural network." In *2018 International Conference on Smart Systems and Technologies (SST)*, pp. 163-167. IEEE, 2018. https://doi.org/10.1109/SST.2018.8564625
[17] Wang, Xing, Tingfa Xu, Jizhou Zhang, Sining Chen, and Yizhou Zhang. "SO-YOLO based WBC detection with Fourier ptychographic microscopy." *IEEE Access* 6 (2018): 51566-51576. https://doi.org/10.1109/ACCESS.2018.2865541
[18] Özyurt, Fatih. "A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine." *Soft Computing* 24, no. 11 (2020): 8163-8172. https://doi.org/10.1007/s00500-019-04383-8
[19] Di Ruberto, Cecilia, Andrea Loddo, and Lorenzo Putzu. "Detection of red and white blood cells from microscopic blood images using a region proposal approach." *Computers in Biology and Medicine* 116 (2020): 103530. https://doi.org/10.1016/j.compbiomed.2019.103530
[20] Wijesinghe, Chinthalanka B., Dilshan N. Wickramarachchi, Iyani N. Kalupahana, R. Lokesha, Indira D. Silva, and Nuwan D. Nanayakkara. "Fully automated detection and classification of white blood cells." In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1816-1819. IEEE, 2020. https://doi.org/10.1109/EMBC44109.2020.9175961
[21] Kutlu, Hüseyin, Engin Avci, and Fatih Özyurt. "White blood cells detection and classification based on regional convolutional neural networks." *Medical Hypotheses* 135 (2020): 109472. https://doi.org/10.1016/j.mehy.2019.109472
[22] Anwar, Shamama, and Afrin Alam. "A convolutional neural network-based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction." *Medical & Biological Engineering & Computing* 58, no. 12 (2020): 3113-3121. https://doi.org/10.1007/s11517-020-02282-x
[23] Saranyan, N., N. Kanthimathi, P. Ramya, N. Kowsalya, and S. Mohanapriya. "Blood Cancer Detection using Machine Learning." In *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1-11. IEEE, 2021. https://doi.org/10.1109/ICECA52323.2021.9675987
[24] Chien, Ju-Huei, Siwa Chan, Shin Cheng, and Yen-Chieh Ouyang. "Identification and detection of immature white blood cells through deep learning." In *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 1-3. IEEE, 2021. https://doi.org/10.1109/LifeTech52111.2021.9391955
[25] Khan, Altaf, Amber Eker, Alexander Chefranov, and Hasan Demirel. "White blood cell type identification using multi-layer convolutional features with an extreme-learning machine." *Biomedical Signal Processing and Control* 69 (2021): 102932. https://doi.org/10.1016/j.bspc.2021.102932
[26] Praveen, Nalla, Narinder Singh Punn, Sanjay Kumar Sonbhadra, Sonali Agarwal, M. Syafrullah, and Krisna Adiyarta. "White blood cell subtype detection and classification." In *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, pp. 203-207. IEEE, 2021. https://doi.org/10.23919/EECSI53397.2021.9624268
[27] Rohaziat, Nurasyeera, Mohd Razali Md Tomari, Wan Nurshazwani Wan Zakaria, and Nurmiza Othman. "White blood cells detection using YOLOv3 with CNN feature extraction models." *International Journal of Advanced Computer Science and Applications* 11, no. 10 (2020). https://doi.org/10.14569/IJACSA.2020.0111058
[28] Liu, Cong, Dengwang Li, and Pu Huang. "ISE-YOLO: Improved squeeze-and-excitation attention module based YOLO for blood cells detection." In *2021 IEEE International Conference on Big Data (Big Data)*, pp. 3911-3916. IEEE, 2021. https://doi.org/10.1109/BigData52589.2021.9672069
[29] Khandekar, Rohan, Prakhya Shastry, Smruthi Jaishankar, Oliver Faust, and Niranjana Sampathila. "Automated blast cell detection for Acute Lymphoblastic Leukemia diagnosis." *Biomedical Signal Processing and Control* 68 (2021): 102690. https://doi.org/10.1016/j.bspc.2021.102690

[30] Shinde, Srushti, Jui Oak, Kajal Shrawagi, and Prachi Mukherji. "Analysis of WBC, RBC, platelets using deep learning." In *2021 IEEE Pune Section International Conference (PuneCon)*, pp. 1-6. IEEE, 2021. https://doi.org/10.1109/PuneCon52575.2021.9686524

[31] Rohaziat, Nurasyeera, Mohd Razali Md Tomari, and Wan Nurshazwani Wan Zakaria. "White blood cells type detection using YOLOv5." In *2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA)*, pp. 1-6. IEEE, 2022. https://doi.org/10.1109/ROMA55875.2022.9915690

[32] Xu, Fanxin, Xiangkui Li, Hang Yang, Yali Wang, and Wei Xiang. "TE-YOLOF: Tiny and efficient YOLOF for blood cell detection." *Biomedical Signal Processing and Control* 73 (2022): 103416. https://doi.org/10.1016/j.bspc.2021.103416

[33] Talukdar, Kabyanil, Kangkana Bora, Lipi B. Mahanta, and Anup K. Das. "A comparative assessment of deep object detection models for blood smear analysis." *Tissue and Cell* 76 (2022): 101761. https://doi.org/10.1016/j.tice.2022.101761

[34] Yentrapragada, Divyateja. "Deep features based convolutional neural network to detect and automatic classification of white blood cells." *Journal of Ambient Intelligence and Humanized Computing* 14, no. 7 (2023): 9191-9205. https://doi.org/10.1007/s12652-022-04422-7

[35] Niu, Zhaoyang, Guoqiang Zhong, and Hui Yu. "A review on the attention mechanism of deep learning." *Neurocomputing* 452 (2021): 48-62. https://doi.org/10.1016/j.neucom.2021.03.091

[36] Lv, Haixin, Jinglong Chen, Tongyang Pan, Tianci Zhang, Yong Feng, and Shen Liu. "Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application." *Measurement* 199 (2022): 111594. https://doi.org/10.1016/j.measurement.2022.111594

[37] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132-7141. 2018. https://doi.org/10.1109/CVPR.2018.00745

[38] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. "Cbam: Convolutional block attention module." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3-19. 2018. https://doi.org/10.1007/978-3-030-01234-2_1

[39] Lee, HyunJae, Hyo-Eun Kim, and Hyeonseob Nam. "Srm: A style-based recalibration module for convolutional neural networks." In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1854-1862. 2019. https://doi.org/10.1109/ICCV.2019.00194

[40] Wang, Chien-Yao, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. "CSPNet: A new backbone that can enhance learning capability of CNN." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 390-391. 2020. https://doi.org/10.1109/CVPRW50498.2020.00203

[41] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, no. 9 (2015): 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

[42] Liu, Shu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. "Path aggregation network for instance segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8759-8768. 2018. https://doi.org/10.1109/CVPR.2018.00913