



Automated counting of white blood cells in thin blood smear images[☆]

Francesca Isabelle F. Escobar ^{a,1}, Jacqueline Rose T. Alipo-on ^{a,1}, Jemima Louise U. Novia ^{a,1}, Myles Joshua T. Tan ^{a,b,*}, Hezerul Abdul Karim ^{c,**}, Nouar AlDahoul ^{c,d,***}

^a Department of Natural Sciences, University of St. La Salle, La Salle Avenue, Bacolod 6100, Philippines

^b Department of Chemical Engineering, University of St. La Salle, La Salle Avenue, Bacolod 6100, Philippines

^c Faculty of Engineering, Multimedia University, Persiaran Multimedia, Cyberjaya, Selangor 63100, Malaysia

^d Computer Science, New York University, Abu Dhabi, United Arab Emirates



ARTICLE INFO

Editor: Huimin Lu

Keywords:

Counting
Detection
Machine learning
White blood cells
YOLOv5

ABSTRACT

Blood cell counting plays a crucial role in clinical diagnosis to evaluate the overall health condition of an individual. Traditionally, blood cells are manually counted using a hemocytometer; however, this task has been found to be time-consuming and error-prone. Recently, machine learning-based approaches have been employed to effectively automate counting tasks. In this work, the fifth version of the 'you only look once' (YOLOv5) object detection method was adopted to automatically detect and count white blood cells (WBCs) in porcine blood smear images. YOLOv5 was chosen because of its speed and accuracy. The dataset used in this study was collected specifically for this WBC counting task. Our experimental results exhibit the high speed and efficiency of YOLOv5 in detecting and counting WBCs, having obtained an accuracy of 89.25% and a mean average precision at 0.5 intersection over union threshold (mAP^{0.5}) of 99%.

1. Introduction

1.1. Background of the study

Blood analysis plays an important role in overall health evaluation for veterinary medicine [1]. The different components of blood provide various kinds of information regarding the physiological state of an individual. A blood smear test is one of the several blood tests used today for both human and veterinary assessment. A thin blood smear test aids in determining the differential blood count and the presence of anomalies in cell morphology, whereas a thick blood smear test aids in blood parasite detection. Blood smear examination is manually performed by observing each sample under a microscope. Such a process is especially time-consuming and labor-intensive, thus posing the risk of error, especially when the analysis of a large number of samples is necessary.

[☆] This paper was recommended for publication by Associate Editor Prof Huimin Lu

* Corresponding author at: Department of Natural Sciences, University of St. La Salle, La Salle Avenue, Bacolod 6100, Philippines.

** Corresponding authors at: Faculty of Engineering, Multimedia University, Persiaran Multimedia, Cyberjaya, Selangor 63100, Malaysia.

E-mail addresses: mj.tan@usls.edu.ph (M.J.T. Tan), hezerul@mmu.edu.my (H. Abdul Karim), nouar.al dahoul@live.iium.edu.my (N. AlDahoul).

¹ These authors have contributed equally to this work.

Artificial intelligence (AI) is currently used to perform certain tasks that originally required humans to carry out. In particular, deep learning (DL), a subset of AI, is a fast-growing and frequent choice for automating various tasks [2]. DL or deep neural networks is a type of neural network composed of a much greater number of processing layers than traditional machine learning algorithms [3,4]. It can learn representations directly from data due to its ability to extract features without manual instruction or preprocessing steps [5]. DL has been shown to yield exceptional results and to surpass conventional techniques. However, creating a good model requires a large dataset and can be computationally expensive. Transfer learning is an approach that involves the use of existing knowledge on a model to learn a new task [6]. With this, training time and the required size of training data is reduced. Several pretrained neural networks have previously been applied for both classification and detection tasks.

In this work, YOLOv5 [7] was employed in the task of detecting and counting white blood cells (WBCs), where the model was retrained for thin blood smear images. A comparison of performance was done for the different variants of the adopted model. This study provides insights into the efficacy of YOLOv5 in detecting blood smear images and the specific version of YOLO that fits the data best.

1.2. Review of related literature

Different approaches have been used in automatic cell counting in various types of data as reported in related literature. Segmentation strategies using image processing techniques have been utilized for cell counting tasks in image data. Dvanesh et al. [8] made use of noise reduction and color channel selection in the quantification of blood cells. Red blood cells (RBCs) were measured through Circular Hough transformation, while boundary detection was applied for WBC counting. Safuan et al. [9] achieved 98% accuracy through color channel subtraction and Circular Hough transform for WBC detection and counting. Another study [10] compared Circular Hough transform and circlet transform for RBC segmentation and counting and found the latter to be superior. While results were promising, they were outperformed by convolutional neural networks (CNNs) [11], a type of deep neural network made of alternating convolution and pooling layers, and completed by a fully connected layer. These networks are commonly used in computer vision [3].

Several studies have utilized various types of CNNs for automatic counting through segmentation [12] and object detection [13,14] based tasks. Segmentation involves pixel-wise masks on the target object whereas object detection uses bounding boxes in locating the target object. Mask R-CNN was used for instance segmentation in the counting and classification of erythrocytes and leukocytes. The study found that it is able to account for overlapping and faded cells, correctly detecting 92% of RBCs and 96% of WBCs. Object detection models also exhibit good results in counting blood cells. RetinaNet [14] was also used in the identification and counting of RBCs, WBCs, and platelets, achieving accuracy values above 97%. Moreover, Faster R-CNN was used in a study by Xia et al. [15] for leukocyte detection, which obtained a detection accuracy of 98%.

You Only Look Once (YOLO) [16] is another popular object detection model used for blood cell images. One study [13] employed the YOLO [16] algorithm for WBCs, RBCs, and platelets with accuracies of 86.89%, 96.09%, and 96.36%, respectively. The study [13] stated that the model performed well but would detect some platelets twice. This challenge was overcome by using k -nearest neighbors (KNN) and the intersection over union (IoU) metric. KNN was applied to find the platelet closest to a detected platelet, then the IoU was used to solve the problem of overlap between two objects. Liu et al. [17] demonstrated high performance in WBC, RBC, and platelet identification with a proposed improvement to the Squeeze-and-Excitation-based YOLO-v3 detection model (ISE-YOLO) involving an improved SE module integrated into YOLOv3. As of this writing, the most recent version of YOLO released is YOLOv5 [7]. Two studies made use of YOLOv5 on the Blood Cell Count and Detection (BCCD) dataset. Shinde et al. [18] attained a mean average precision of 93%, yielding performance superior to previous versions of YOLO. Moreover, Zhao et al. [19] were able to achieve a mean average precision of 96%, demonstrating superiority over SSD-MobileNet.

This paper is organized as follows: Section 2 describes the datasets and the model implementation. In Section 3, the experimental results are discussed in detail. Finally, Section 4 summarizes the outcome and significance of this work, and gives readers a glimpse into potential improvements to our work in the future.

2. Materials and methods

2.1. Dataset

We prepared the dataset [20] used for this study under the supervision of a licensed veterinarian. Peripheral blood films were prepared with blood samples collected from juvenile pigs and were fixed with Hema-Quick (Pappen-heim) staining. Blood collection was performed by a veterinarian while blood smearing and staining were carried out by us. Data acquisition involved capturing the digital images of the blood smear under a microscope. The setup involved a smartphone mounted on an Olympus® CX23 compound microscope set under oil immersion magnification.

The dataset generated consisted of 655 images with dimensions 3150×3150 pixels, which were resized into 416×416 pixels. Images were annotated by bounding boxes around the WBCs present in each image. Image labels were saved as .xml and .txt files. Label files follow the YOLO [16] format, containing the label of the WBC in the `<name>` tag and its corresponding location which was represented by `<xmin>`, `<ymin>`, `<xmax>`, `<ymax>` tags. The dataset was converted to YOLOv5 format, where the annotations were in PyTorch TXT Annotation Format. The dataset was then split into a 70:15:15 ratio for training, validation, and testing.

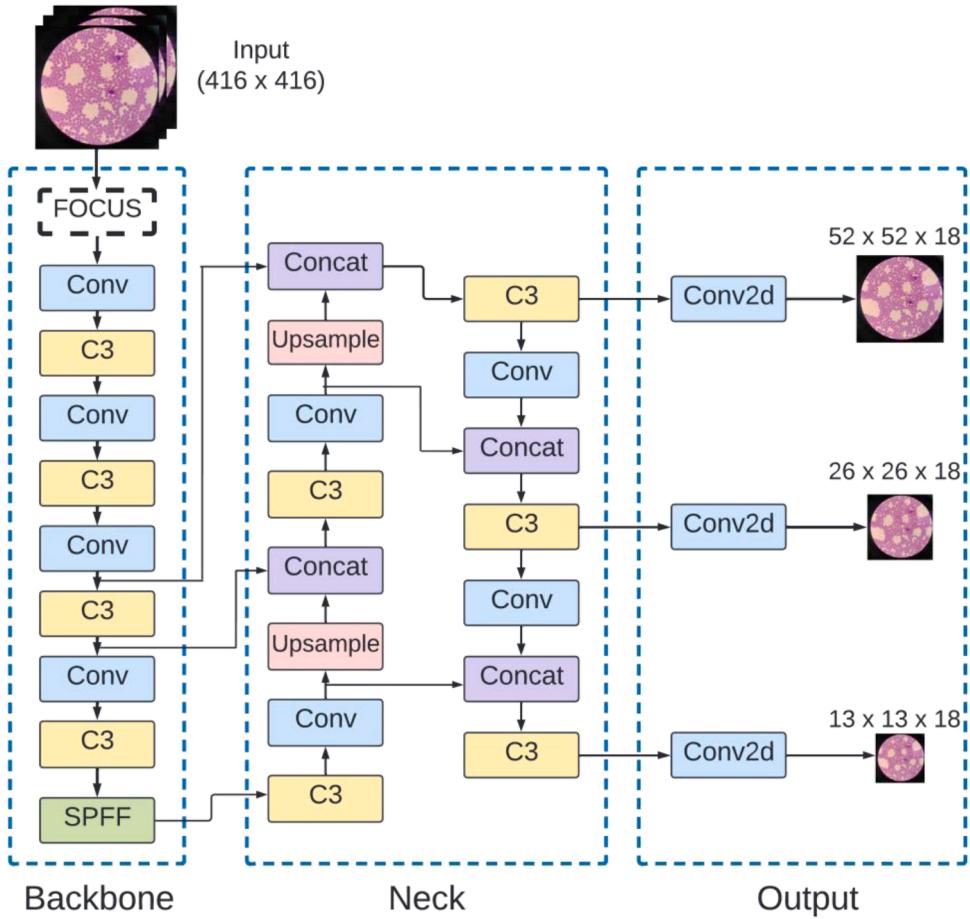


Fig. 1. The general architecture of YOLOv5 [7] used for detection and counting, partly based on Nepal & Eslamiat [22].

2.2. Model implementation

The adopted machine learning algorithm in this study was YOLOv5 [7], the fifth generation of YOLO [16] that uses the PyTorch™ framework and utilizes Hard Swish as its activation function. This state-of-the-art algorithm was developed by Ultralytics and released via a repository on GitHub. YOLOv5 demonstrated better results when applied on blood cell detection than older YOLO model versions [18]. The general architecture mainly consists of a backbone, a neck and a head or the output as shown in Fig. 1. The cross stage partial network (CSPnet) that was incorporated into Darknet serves as the backbone, the path aggregation network (PANet) serves as the neck, and YOLO [16] serves as the head. The backbone is responsible for feature extraction; the neck for the aggregation of features; and the head for the target predictions and outputs.

The framework comes with five varieties differing in size and computational cost. These include YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, which were pretrained on the COCO dataset [21]. To determine which specific variant suited the model, all variants were applied on the dataset for comparison. The task was restricted to one class due to limitations in the dataset, where a large imbalance among WBC types was observed. All variants were trained with 70 epochs as they required fewer iterations. The batch size was set to eight images to match the GPU resources.

For counting, the model weights from the training were saved and used to make inferences on the testing set. Three different confidence thresholds were selected: 0.4, 0.5, and 0.6. The confidence threshold is an important parameter in the YOLO [16] algorithm because it makes use of a threshold to predict each grid cell.

All experiments were executed in Colaboratory™ by Google, an online Jupyter™ notebook-based environment that provides access to computing services covering central processing unit (CPU) and graphics processing unit (GPU) resources. We used 25GB RAM and an NVIDIA® Tesla® K80 GPU.

2.3. Performance evaluation

The parameters used to evaluate the performance of the model included accuracy, precision, recall, mean average precision, and mean absolute error. The definition of these metrics is described below:

Table 1

Evaluation Metrics for Five Variants of YOLOv5.

	Precision	Recall	mAP (0.5 IoU)	mAP (0.5 to 0.95, step 0.05 IoU)
YOLOv5n	0.97	0.93	0.98	0.70
YOLOv5s	0.99	0.98	0.99	0.82
YOLOv5m	0.98	0.94	0.98	0.77
YOLOv5l	0.97	0.95	0.99	0.77
YOLOv5x	0.99	0.91	0.99	0.78

Precision [23] refers to the proportion of the number of white blood cells correctly predicted and regarded as true positives (TP) to the total number of true positives and false positives (FP) which are non-WBC that were detected as WBC:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

Recall [23] refers to the proportion of the number of correctly predicted cells to the sum of the number of ground truths that includes the TP and false negatives (FN):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

The mean average precision (mAP) [23] refers to the evaluation of the model's results by comparing the bounding box in the ground-truth image versus the detected box. The mAP computes the mean average precision (AP) score of the model across the different classes.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (3)$$

Where AP_i refers to the average precision of class i , and N represents the number of classes.

Average ground truth (AGT) describes the mean WBC count found in each image of the testing set.

$$\text{AGT} = \frac{\text{SumofWBCsperimage}}{\text{Totalnumberoftestimages}} \quad (4)$$

Ground truth standard deviation (GTSD) measures the degree of variation within the WBC count of each sample of the test dataset.

$$\delta_{\text{GTSD}} = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \quad (5)$$

Where x_i refers to the count of WBCs per image, μ refers to the mean of ground truth WBC count, and N represents the number of samples. Average estimated number (AEN) describes the mean WBC count predicted by the model in each sample of the test set.

$$\text{AEN} = \frac{\text{Sum of predicted WBCs per image}}{\text{Total number of test images}} \quad (6)$$

Estimated number standard deviation (ENSD) measures the degree of variation within the WBC predicted count of each sample of the test dataset.

$$\delta_{\text{ENSD}} = \sqrt{\frac{\sum (y_i - \mu)^2}{N}} \quad (7)$$

Where y_i refers to the predicted count of WBCs per image, μ refers to the mean of predicted WBC count, and N represents the number of samples.

The MAE is used to compute the errors between the actual numbers and the number predicted by the model. This is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^N |y - \hat{y}| \quad (8)$$

Where N is the total number of observations, y is the actual value, and \hat{y} is the predicted value.

Accuracy refers to the performance of the model based on the comparison of the predicted cell count of the model with the ground truth or the actual count:

$$pa(\%) = \left[1 - \frac{|pc - ac|}{ac} \right] \times 100 \quad (9)$$

Where accuracy expressed in percentage is denoted by pa , predicted count by pc , and actual count by ac . This metric was used in each image of the testing set, and the average accuracy was obtained from the accuracies of all images.

Table 2

Counting Results of Five Variants of YOLOv5 for Three Different Confidence Thresholds.

Model variant	YOLOv5n			YOLOv5s			YOLOv5m			YOLOv5l			YOLOv5x		
Conf. threshold	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.6
AGT	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
GTSd	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
AEN	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
ENSD	1.21	1.17	1.15	0.19	0.19	0.19	0.22	0.27	0.26	0.22	0.27	0.26	0.22	0.27	0.26
MAE	0.22	0.27	0.26	0.19	0.19	0.19	0.22	0.27	0.26	0.22	0.27	0.26	0.22	0.27	0.26
pa	84.66	86.19	86.87	89.25	89.25	89.25	88.83	87.55	86.02	87.33	87.33	86.05	88.71	89.02	86.36

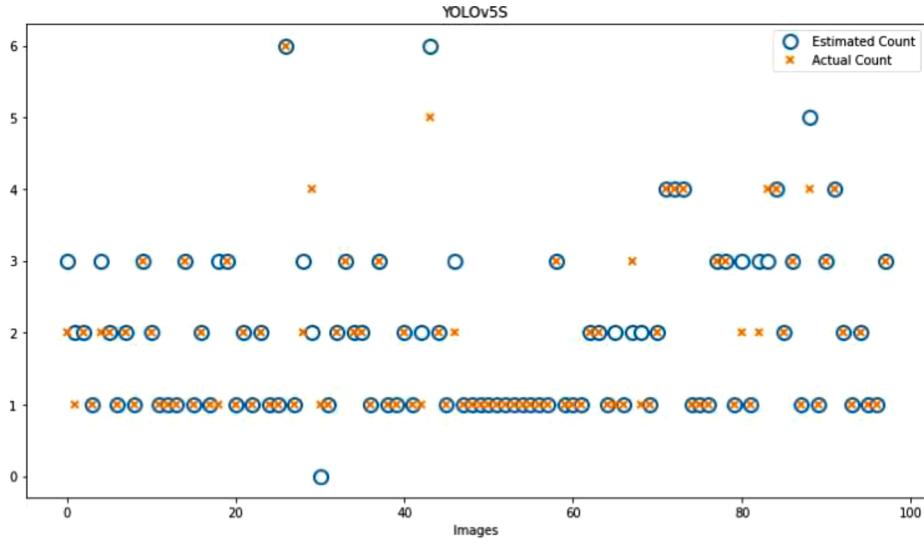


Fig. 2. The estimated and actual wbc count for each image using YOLOv_{5S}.

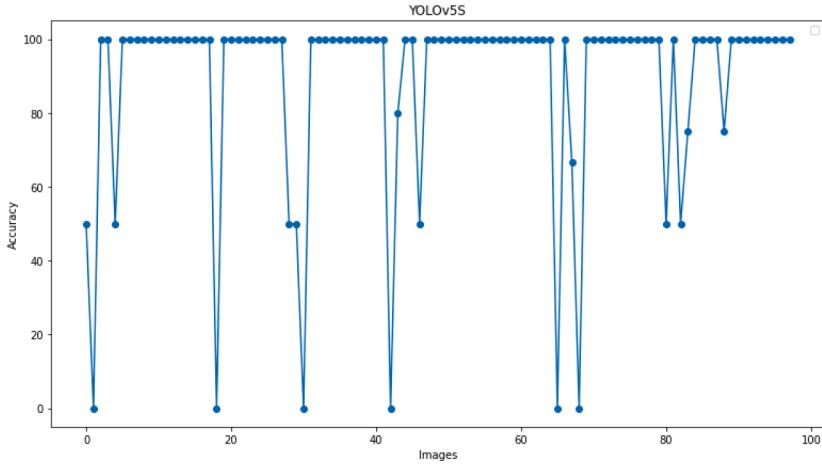


Fig. 3. The accuracy (pa%) of WBC counting for each image using YOLOv5s.

Difference refers to the performance of the model based on the difference of the predicted cell count of the model with the ground truth or the actual count:

$$\text{Difference (\%)} = \frac{|pc - ac|}{\frac{|pc+ac|}{2}} \times 100 \quad (10)$$

Where predicted count is denoted by pc and actual count by ac . This metric was used in each image of the testing set.

3. Results and discussion

To determine which variant offers the best performance and accuracy in terms of detection and counting, five variants of YOLOv5, differing in size and computational cost, were compared.

Table 1 shows the performance of the five variants through evaluation metrics. Among the variants, YOLOv5s yielded the highest model evaluation metrics (precision and recall). YOLOv5s has 213 layers with seven million parameters. It also outperformed other models yielding the highest mAP^{0.5} and mAP^{0.5:0.95}. This means that the WBCs detected were well localized by the bounding boxes. Usually, larger models yield better mAP levels and accuracy, so one might expect that YOLOv5x would yield the best results. However, in this study, the model was outperformed by YOLOv5s. This result agrees with one study [24] whose findings suggest that a shallow network would better suit the training of small datasets.

In Table 2, three different thresholds were applied in evaluating the performance of different YOLOv5 variants in counting using

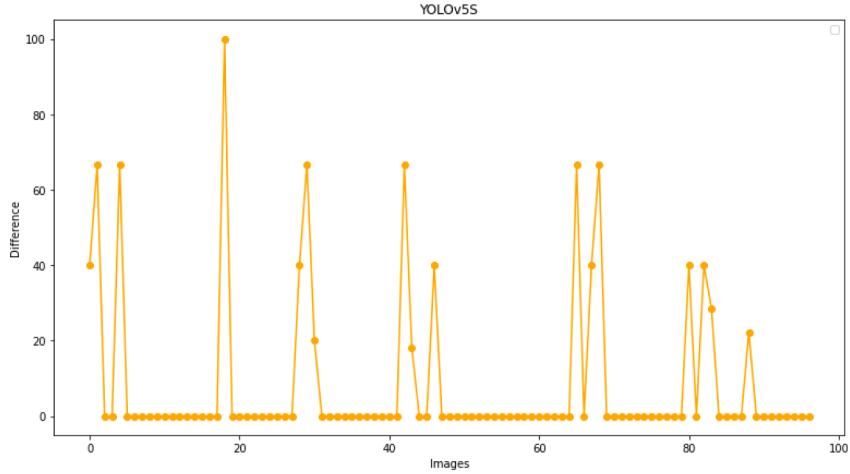


Fig. 4. The difference (%) in WBC count for each image using YOLOv5s.

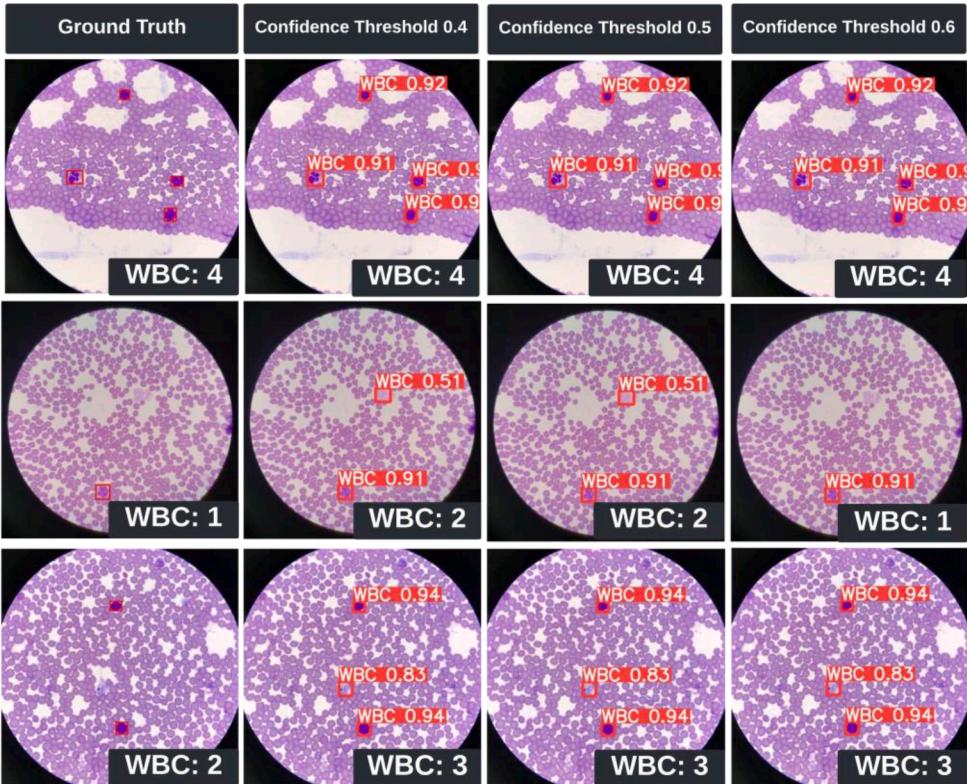


Fig. 5. Image samples with WBCs detected using YOLOv5s across confidence thresholds: 0.4, 0.5, and 0.6.

the same testing set. Note that the following acronyms correspond to the respective terms: average ground truth (AGT) found in all images, ground truth standard deviation (GTSD), average estimated number (AEN) predicted by the models, estimated number standard deviation (ENSD), mean absolute error (MAE), and accuracy in percentage (pa). The average and estimated ground truths were rounded off. The average accuracy and the MAE show that YOLOv5s performed best in counting WBC at a confidence threshold of 0.5. This is consistent with its detection performance in Table 1. It can be observed that most models with a confidence threshold of 0.4 have higher accuracies than those with other thresholds. It was mentioned that setting a higher threshold may cause the model to become robust to boxes that contain an object. Thus, fewer positive predictions would be made and consequently, would lead to an increase in false negatives and a decrease in false positives [25].

Fig. 2 shows the performance of the model that yielded the highest accuracy, YOLOv5s at thresholds of 0.4 and 0.5, based on

estimated and actual count results. These two thresholds were used since they presented the same results in terms of the evaluation metrics and the number of cells they estimated in each image. Because the dataset was limited in terms of variation, with most images containing only two instances of WBCs, cells that were undetected or subjects that were misidentified to be WBCs had a large impact on the measurement of the model's accuracy. Nevertheless, it is important to note that the differences between the estimated and the actual counts were minimal.

Another factor that may have affected the average accuracy of the models are the blob-shaped residues that could easily be mistaken for cells. As seen in Fig. 2, most images have only one WBC, but most models estimated the number of cells to be two or more, because the residues present in the image were misidentified as WBCs.

Fig. 3 shows the accuracy of the prediction of WBC count in each image and the number of underestimated or overestimated cells expressed as a difference in percentage. As most images contain only one or two instances of WBC, an error in prediction would drastically affect the accuracy of the model. Thus, some images have accuracies that dip to 0% and differences as high as 100%.

Fig. 4 is a graphical representation of the difference in WBC count between the ground truth and predicted value in each image. Like Fig. 3, the measured difference would appear large for some samples because of the low WBC count in the ground truth.

Fig. 5 provides image samples of WBCs detected by YOLOv5s across different confidence thresholds. Numbers indicated on the right bottom corner correspond to the total WBC count per image. Values set along the bounding boxes are the confidence scores measured by the model. At lower thresholds, certain objects are misidentified to be leukocytes. Artifacts from smearing and staining may also give rise to false detection as observed in the third row of images, where three, instead of two, cells are detected.

4. Conclusion

In summary, we demonstrated a new application of YOLOv₅ through this work. This task includes automatic detection and counting of WBCs in thin blood smear images of Visayan warty pigs. YOLOv5s was selected because it was able to balance high speed with superior performance. YOLOv5s, which is the second smallest YOLO variant was found to outperform other variants and showed that optimal model size usually depends on the dataset size. The detection task yielded the following performance metrics: 99% mAP^{0.5}, and 82% mAP^{0.5:0.95}. The results produced by our experiments confirmed the advantage of using YOLO for faster laboratory testing to count WBCs. In the future, further studies can be done to explore the use of models for detection and counting of RBCs and platelets. Additionally, experiments to explore preprocessing methods to overcome the challenges brought about by stain-associated artifacts can be carried out. Furthermore, we intend to increase the size of our training data to improve detection and counting performance.

CRediT authorship contribution statement

Francesca Isabelle F. Escobar: Conceptualization, Software, Investigation, Writing – original draft. **Jacqueline Rose T. Alipon:** Conceptualization, Software, Investigation, Writing – original draft. **Jemima Louise U. Novia:** Conceptualization, Software, Investigation, Writing – original draft. **Myles Joshua T. Tan:** Supervision, Writing – review & editing. **Hezerul Abdul Karim:** Project administration, Funding acquisition. **Nouar Aldahoul:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] Jezek J, Staric J, Nemec M, Plut J, Golinar Oven I, Klinkon M, et al. The influence of age, farm, and physiological status on pig hematological profiles. *J Swine Health Prod* 2018;26:72–8. 03.
- [2] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 2021;8(1).
- [3] Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. *IEEE Access* 2019;7:53040–65.
- [4] Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics* 2019;8(3):292. Basel.
- [5] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [6] Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010;22(10):1345–59.
- [7] Jocher G., Chaurasia A., Stoken A., Borovec J., NanoCode012, Kwon Y, et al.. Ultralytics/YOLOv5: v6.1 – Tensorrt, tensorflow edge TPU and openvino export and inference. Zenodo; 2022. Available.
- [8] Dvanesh VD, Lakshmi PS, Reddy K, Vasavi AS. blood cell count using digital image processing. In: Proceedings of the international conference on current trends towards converging technologies (ICCTCT); 2018. p. 1–7.
- [9] Safuan SN, Tomari R, Zakaria WN, Othman N. White blood cell counting analysis of blood smear images using various segmentation strategies. In: Proceedings of the AIP Conference Proceedings; 2017.
- [10] Sarrafzadeh O, Dehnavi AM, Rabbani H, Ghane N, Talebi A. Circlet based framework for red blood cells segmentation and counting. In: Proceedings of the IEEE workshop on signal processing systems (SiPS); 2015. p. 1–6.

- [11] Lavitt F, Rijlaarsdam DJ, van der Linden D, Weglarz-Tomczak E, Tomczak JM. Deep learning and transfer learning for automatic cell counting in microscope images of human cancer cell lines. *Appl Sci* 2021;11(11):4912.
- [12] Dhib N, Ghazzai H, Besbes H, Massoud Y. An automated blood cells counting and classification framework using mask R-CNN deep learning model. In: Proceedings of the 31st international conference on microelectronics (ICM); 2019. p. 300–3.
- [13] Alam MM, Islam MT. Machine learning approach of automatic identification and counting of blood cells. *Healthc Technol Lett* 2019;6(4):103–8.
- [14] Dra-lus G Mazur D, Czmil A. Automatic detection and counting of blood cells in smear images using retinanet. *Entropy* 2021;23(11):1522. <https://doi.org/10.3390/entropy23111522>.
- [15] Xia T, Jiang R, Fu YQ, Jin N. Automated blood cell detection and counting via deep learning for microfluidic point-of-care medical devices. *IOP Conf Ser Mater Sci Eng* 2019;646(1):012048.
- [16] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Las Vegas, NV, USA: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 779–88.
- [17] Liu C, Li D, Huang P. ISE-YOLO: Improved squeeze-and-excitation attention module based YOLO for blood cells detection. In: Proceedings of the IEEE international conference on big data (Big Data); 2021. p. 3911–6.
- [18] Shinde S, Oak J, Shrawagi K, Mukherji P. Analysis of WBC, RBC, platelets using deep learning. In: Proceedings of the IEEE pune section international conference; 2021. p. 1–6.
- [19] Zhao J, Cheng Y, Ma X. A real time intelligent detection and counting method of cells based on YOLOv5. In: Proceedings of the IEEE international conference on electrical engineering, big data and algorithms (EEBDA); 2022. p. 675–9.
- [20] Alipo-on J.R., Escobar F.I., Novia J.L., Atienza M.M., Tan M.J., Alda-houl N. Dataset for detection of white blood cells of the juvenile visayan warty pig in blood smear images; 2022. Available from: <https://github.com/FrancescaIE/WBC-detection-dataset.git>.
- [21] Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. Lecture Notes in Computer Science, 8693. Cham: Springer; 2014.
- [22] Nepal U, Eslamiat H. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVS. *Sensors* 2022;22(2):464.
- [23] Padilla R, Netto SL, da Silva EAB. A survey on performance metrics for object-detection algorithms. In: Proceedings of the international conference on systems, signals and image processing (IWSSIP); 2020. p. 237–42.
- [24] Nguyen ND, Do T, Ngo TD, Le DD. An evaluation of deep learning methods for small object detection. *J Electr Comput Eng* 2020;2020:1–18.
- [25] Wenkel S, Alhazmi K, Liiv T, Alrshoud S, Simon M. Confidence score: the forgotten dimension of object detection performance. *Sensors* 2021;21(13):4350.

FRANCESCA ISABELLE FLORES ESCOBAR is a graduate at the University of St. La Salle, Bacolod, Philippines with a degree in Bachelor of Science in Biology in 2022. Her research interests include computer vision, deep learning, and natural sciences.

JACQUELINE ROSE T. ALIPO-ON is a graduate of Bachelor of Science in Biology and is currently a medical student at the University of St. La Salle in Bacolod City, Philippines. Her research interests cover natural sciences, image processing, and deep learning.

JEMIMA LOUISE U. NOVIA is a student of the College of Medicine at the University of St. La Salle, Bacolod City, Philippines. She has received a B.S. degree in biology from the same university in 2022. Her interests include natural sciences, deep learning, and computer vision.

MYLES JOSHUA TOLEDO TAN (Member, IEEE) received the B.S. degree, summa cum laude in biomedical engineering from the University at Buffalo, The State University of New York, in 2017, and the M.S. degree in applied biomedical engineering from Johns Hopkins University, USA, in 2018. He has been an Assistant Professor of natural sciences and chemical engineering with the University of St. La Salle (USLS), since 2018.

HEZERUL ABDUL KARIM (Senior Member, IEEE) received the M.Eng. degree in science from Multimedia University, Malaysia, in 2003, and the Ph.D. degree from the University of Surrey, U.K., in 2008. He is currently a Professor with the Faculty of Engineering, Multimedia University. He is currently serving as a Chair in the IEEE Signal Processing Society Malaysia Chapter.

NOUAR ALDAHOUL received the B.Eng. and M.Eng. degrees in computer engineering from Damascus University, in 2008 and 2012, respectively, and the Ph.D. degree in machine learning from International Islamic University Malaysia, in 2019. She is currently a postdoctoral associate in New York University, Abu Dhabi, UAE. Her main research interests include deep learning, computer vision, and Natural Language Processing.