Deep Learning approach to Predict blood cell and counting

BY

MD ABDUR RAHAMAN
ID 171-35-194

A thesis submitted fin partial fulfillment of the requirement for the degree of Bachelor of Science in Software Engineering.

Department of Software Engineering
DAFFODIL INTERNATIONAL UNIVERSITY
Semester Summer – Year-2021

IM: Dr. Imran Mahmud; SMR: S A M Matiur Rahman; RZ: Raihana Zannat; NH: Nayeem Hasan; SI: Mr. Shariful Islam; SFR: SK. Fazlee Rabby; MA: MarziaAhmed; RM: Md. Rajib Mia.

THESIS DECLARATION

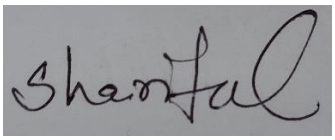
I'm, Md. Abdur Rahaman, hereby declare that this report was written under the supervision of Shariful Islam, Lecturer, Daffodil International University Department of Software Engineering. We further certify that neither this study nor any component of it has been submitted for a degree elsewhere.

Supervised By,
Shariful Islam
Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Submitted By,
Md. Abdur Rahaman
ID: 171-35-194
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Approval

This **Thesis** is titled **"Deep Learning approach to Predict blood cell and counting",** submitted by **MD. Abdur Rahaman, ID: 171-35-194** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.



Shariful Islam
Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner

ACKNOWLEDGEMENT

First, I am expressing my gratitude to the almighty Allah for giving me the ability to complete this thesis work. I would like to express our sincere gratitude to my honorable supervisor, Shariful Islam, Lecturer, Department of Software Engineering. This thesis would not have been completed without his support and guidance. His constant encouragement gave me the confidence to carry out my work. I would also like to give special gratitude to one of my favorite teachers Rajib Mia, Lecturer, Dept. of Software Engineering, Daffodil International University. His proper direction and guidance help me to prepare this thesis work without any difficulty. I express my heartiest gratitude towards the entire department of Software Engineering at Daffodil International University for providing good education and knowledge. I also express my gratitude to all my teacher's SAM Matiur Rahman, Associate Professor; Dr. Imran Mahmud, Associate Professor and Head, Dept. of Software Engineering. I would like to convey my sincere gratitude to Bikash Kumar Paul, Lecturer, Dept. of Information and Communication Technology, Mawlana Bhashani Science and Technology University. The knowledge that I have learned from the classes in my degree of bachelor's in software engineering level were essential for this thesis. In course of conducting the study necessary information were collected through books, journals, electronic media and other secondary sources. I also want to thank my friends for providing me support and encouragement. Their optimism and encouragement have allowed to overcome any obstacle at any phase.

ABSTRACT

The counting of blood cells is an important test that aids in the diagnosis of certain disorders. The technique of manually counting abnormal blood cells by an experienced specialist is exceedingly arduous, time-consuming, and imprecise, with a substantial risk of error. Automated detection of blood cells using image processing techniques is gaining prominence as a result of recent advancements. In this research, the authors offer a deep learning strategy for automatically identifying and counting three types of blood cells using the ‘you only look once' (YOLO) object detection and classification algorithm in this paper. The YOLO framework has been taught to automatically recognize and count red blood cells, white blood cells, and platelets using a modified configuration BCCD Dataset of blood smear image. Overall, our automated blood cells counting system is fast and more efficient to detect blood cells.

TABLE OF CONTENT

Contents Page

LIST OF FIGURES

CHAPTER: 1 INTRODUCTION

1.1 Background:

A complete blood cell (CBC) count is a vital test that medical practitioners frequently order to assess a patient's health [1, 2]. Red blood cells (RBCs), white blood cells (WBCs), and platelets are the three primary types of cells that make up blood. RBCs, also known as erythrocytes, make up 40–45 percent of all blood cells [American Society of Haematology: http://www.hematology.org/Patients/Basics/]. Blood contains a large number of platelets, also known as thrombocytes. WBCs, commonly known as leukocytes, account for less than 1% of all blood cells. RBCs transport oxygen to our body tissues, and the number of RBCs determines the amount of oxygen delivered to those tissues. WBCs combat infections, while platelets aid in blood coagulation. Because these blood cells are so numerous, typical manual blood cell counting with a haemocytometer is both time consuming and inaccurate, and accuracy rests heavily on the clinical laboratory analyst's abilities [3, 4]. As a result, an automated procedure for counting distinct blood cells from a smear image will make the entire counting procedure much easier. Image categorization and object recognition applications are getting more robust and accurate as machine learning techniques advance. As a result, technologies based on machine learning are being used in a variety of sectors. Deep learning methods are being used in a variety of medical applications, including anomaly identification and localization in chest X-rays [5], automatic left ventricle segmentation in cardiac MRI [6], and diabetic retinopathy identification in retinal fundus pictures [7]. Thus, it is worthwhile to investigate deep learning-based approaches for identifying and counting blood cells in smear images. A blood cell counting method based on deep learning is presented in this Letter. It has been suggested. We use a deep learning-based object recognition system. A method for detecting distinct types of blood cells. Among them are object detection methods that are state-of-the-art, such as regions with You only have to look once YOLO) with a convolutional neural network (R-CNN) [8]. From blood smear images, we retrain the YOLO framework to automatically detect and count RBCs, WBCs, and platelets.

A verification mechanism has been devised to avoid the framework's repeated counting in order to increase counting accuracy. In addition, the trained model was evaluated using photos from a different dataset to see if the strategy was generalizable. The proposed deep learning-based blood cell identification and counting system is shown in Figure

1.2 Motivation of Research

Human blood is made up of three types of cells: red blood cells (RBCs), white blood cells (WBCs), and platelets. The most abundant cell type in human blood, red blood cells (RBCs) transport oxygen from the lungs to all tissues and cells, as well as carbon dioxide from the tissues back to the lungs. Low RBC counts in the blood can cause a variety of pathophysiological issues. Hemolytic anemia, sickle cell anemia, thallasemia, and different membranopathies and enzymopathies can all be diagnosed, evaluated, and monitored using RBC counting. As a result, analyzing microscopic pictures is critical for determining the typical form and size of RBC.
Blood counts were done manually in the past pathological method, which was more error-prone and time-consuming. Because of its low cost and ease of deployment, automatic RBC detection has recently acquired a lot of traction. The cells' chemical and physical properties are determined. It just requires a minimal bit of specimen and properly counts the cells, but it fails to detect abnormalities and is quite expensive. The goal of this study is to create a system that can reliably count red blood cells and identify the quantity of normal and pathological cells in a picture. In the field of medicine, the analysis of blood smear images is critical since it aids in the early diagnosis of disease. It helps clinicians make the best decisions about illness monitoring and, as a result, disease prevention. Separating red blood cells from white blood cells is the first step in detecting them. If white blood cells aren't separated, they can be mistaken for red blood cells. The labeling method and circular Hough transform are used to create a counter.

1.3 Problem Statement:

Since learning about previous similar works, I've noticed some weaknesses, which may take this to the next level of the study.

✓ Limited label images.

✓ XML and Images are in same folder.

✓ Worst images are there.

1.4 Research Question:

A research question is an answerable inquiry into a specific problem or topic. In a research, this is the initial step. Once, an understanding of what to explore, the 'initial step' means that the research topic is the first important step in the research. Some questions are explored. The list of questions is given below:

✓ How to collect the data set for selected diseases? and where?

✓ How to data mine the data set?

✓ How to process data?

✓ What was the splitting ratio?

✓ Which algorithm use there?

✓ Why use this algorithm?

✓ How to add the references?

1.5 Research Objectives:

The key goals of this thesis are given below:

✓ Finding out how well yolov5 can detect blood cells from better images and worse images.

✓ How fast YOLOv5s algorithm for detecting blood cells.

✓ Detect and Count blood cells.

✓ Why YOLOv5s better than others algorithms.

1.6 Research Scope:

The study's scope applies to the areas covered by the research. The scopes are given below:

✓ Downloading data.

✓ Data Processing using Python.

✓ Apply Algorithm.

✓ Training of data.

✓ Detection and Counting blood cells.

1.7 Thesis Organization:

In a certain section, the whole paper is organized. Where the relevant analysis is discussed in Section 2 to extract the conceptual framework. The source of supporting proof of arguments to not only validate the statement or theory, but also serve as a basis for the results. The Proposed methodology is presented in the section 3. I have clarified my results and final outcome in section 4 with the support of some graphs and network. In that section, all the outcomes have been described. A small summery of the whole research is presented as conclusion in section 5

2.1 Definition of The Thesis Topic or Area:

In general, when it comes to automated blood cell counting, there are two basic ways. They are the methods for image processing as well as the machine learning strategy

2.2 The Direction of Research:

Our goal is to recognize and count blood cells directly from a smear image using the YOLO object recognition and classification technique. With an updated configuration and annotated blood cells training image, we must train the YOLO framework.

2.3 Theoretical Framework:

YOLO (You Only Look Once) is a cutting-edge object detection categorization method. Object detection is treated as a regression problem. To make a quick forecast for both image class and location, only one forward propagation pass over the network is required. It resizes the image by 448 448 pixels and divides it into 7 7 grid cells, with each grid cell predicting two bounding boxes and their confidence scores. If the object's center falls inside a grid cell, that grid cell is in charge of detecting the object.

2.4 Resources:

To recognize automobiles in this research, we employed the most recent version of the You Only Look Once (YOLO) object detection algorithm. Our concentration as computer vision practitioners was on the algorithm's application, data collecting, and data annotation.

Figure 1 represent the method step by step

Figure 1

3.1 Data Collection:

We use a publicly available dataset of annotated blood cell images called Blood Cell Count Dataset (BCCD) [BCCD:https://github.com/akshaylamba/all_CELL_data].

3.2 Preprocessing:

There are 364 annotated smear photos in all, but the dataset includes a major defect. We discovered that one annotation file in the test set does not include any RBCs, despite the image having RBCs, after splitting the dataset into training (300) and testing (64). Furthermore, three annotated files have significantly lower RBC than the actual RBC. As a result, four erroneous files are removed, reducing the overall size of the test set to 60. We chose 60 training photos with annotations at random for the validation set.

3.3 YOLOv5s:

YOLO (You Only Look Once) is a cutting-edge object detection categorization method. Object detection is treated as a regression problem. It only necessitates to build a quick connection, one forward propagation pass through the network is required. Both image class and location are predicted. The image is resized by 448 448 and divides the entire image into a grid cell of 7 7 pixels and each grid cell foresees two boundary boxes The boxes have a confidence score. If the object's center falls, enters a grid cell, and the grid cell is in charge of detecting it that particular object The YOLO approach was first implemented as a On the PASCAL VOC dataset, a CNN was assessed. Figure 2 show that YOLO architecture.

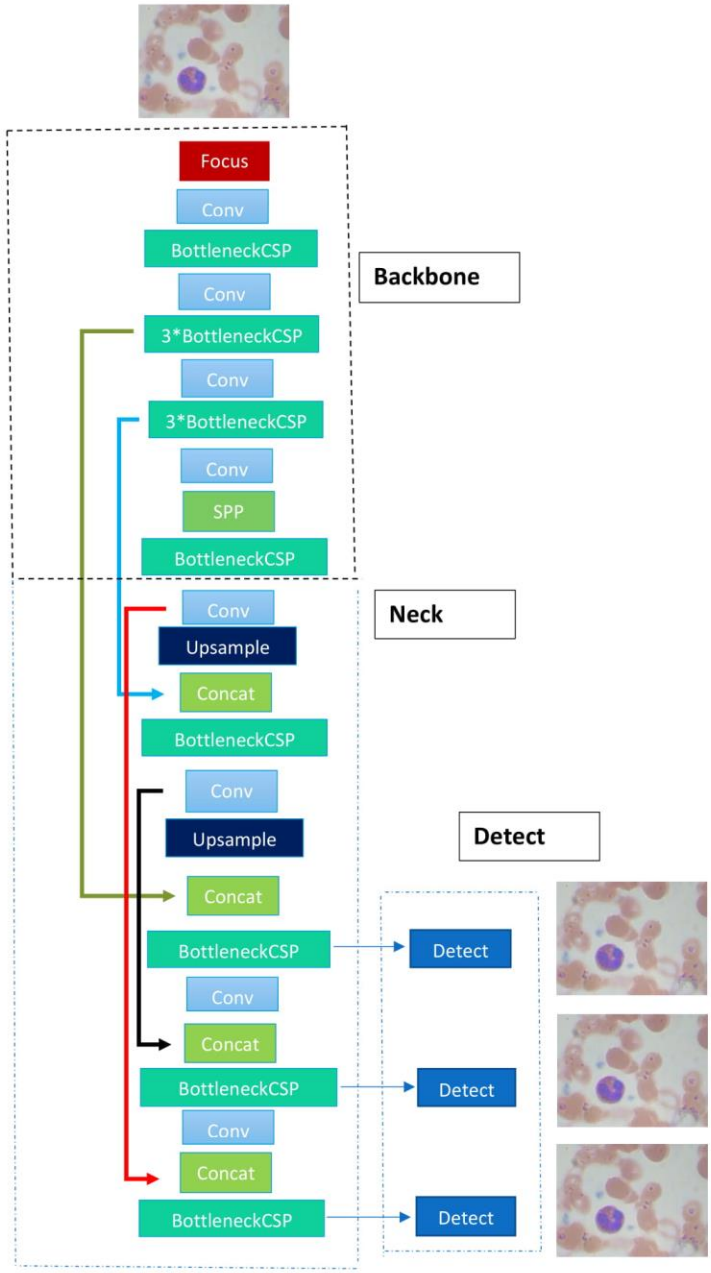


Figure.2 Yolov5s Architecture

3.4. Training Of data:

The Tiny YOLO configuration was trained for 20 different classes when it was first implemented. We adjust it for three kinds of blood cells: WBC, RBC, and platelets, in order to use it for blood cell identification. The number of filters in the final convolutional layer of the CNN architecture must also be modified as a result of changing the class number. For each anchor box, YOLO forecasts five values as well as class probabilities. The values are the probability of having an object in a grid cell, the object's x and y coordinates, height, and width. The number of anchor boxes in our case is 5, which gives us more flexibility in placing bounding boxes based on the object's aspect ratio.

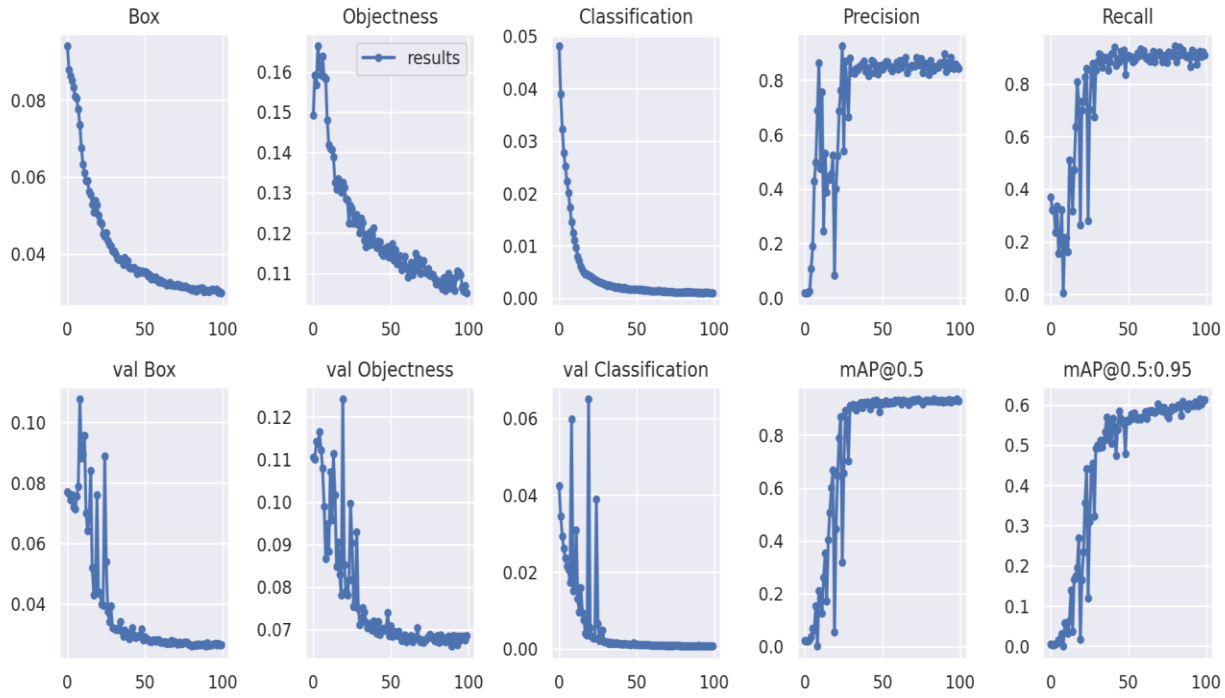


Figure.3 Plot graph shown

3.4. Proposed identification and counting method of blood cell:

Our suggested method is a machine learning strategy in which we employ the YOLO algorithm to identify and count blood cells automatically. It contains a training model with a modified configuration in which the final convolution layer is changed for three outputs, and blood cell identification using an appropriate algorithm. Cells are counted based on their labels. Depending on the detected cell, the modified YOLO returns three labels: RBC, WBC, and Platelets. The total number of RBC in a smear image is equal to the total number of labels containing the word "RBC," the total number of WBC is equal to the total number of labels containing the word "WBC," and so on. Our models provide two separate detections for a single platelet in some circumstances. The reason for this is that the same platelet is detected in two successive grid cells, resulting in the same platelet being tallied twice. To avoid double counting, we use the KNN method in each platelet to determine its closest platelet, and then we compute the extent of overlap between two platelets using the intersection of union (IOU). We allow 10% overlap between platelets and their closest platelets based on empirical data. If the overlap is greater than that, we treat the cell as a double count in order to avoid erroneous counting.

RBCs, WBCs, and platelets are automatically identified and counted using the proposed method. We put our model to the test on a test dataset of 60 photos with known ground facts. First, we count the distinct cells in the validation dataset with different confidence thresholds using our model. It should be noted that the threshold is important in YOLO because it is used to anticipate each grid cell rather than the entire image. Grid cells with no blood cells have a low confidence level. By selecting a suitable confidence threshold, we can eliminate duplicate and erroneous predictions. We calculate the average absolute error between ground truths and the validation dataset's predicted number of cells. We get the minimum average by using different confidence thresholds. Choose those confidence values in the blood cell identification process based on the absolute error value for each type of cell. where cell is the cell type (RBC, WBC, or platelets), N is the size of the validation dataset (in our case, 60), x is the number of cells, and 1 denotes the average absolute error value for that cell. Table 1 shows the calculated error values.
As can be seen from the table, a nominal threshold of 0.55 can be used to count RBCs. However, the threshold for WBC and platelets is substantially lower (0.35 and 0.25 in our trials). As a result, the suitable thresholds for each cell type are chosen as follows:

	RBC	**WBC**	**Platelets**
Estimated	790	70	82
Accuracy	86%	93%	80%

Then, using the total number of ground truth cells and the total number of estimated cells in the test dataset, we determined the accuracy. We achieved 83 percent accuracy for RBC with a confidence threshold of 55 percent. Table 2 shows the total estimated numbers of cells of various categories with accuracy determined at their respective confidence threshold value. It may appear that the proposed method is counting RBCs that aren't visible in the photos. However, we should point out that the ground truth labels were missing for several of the RBCs along the image's periphery. The YOLO algorithm can detect these RBCs, resulting in a high RBC count.

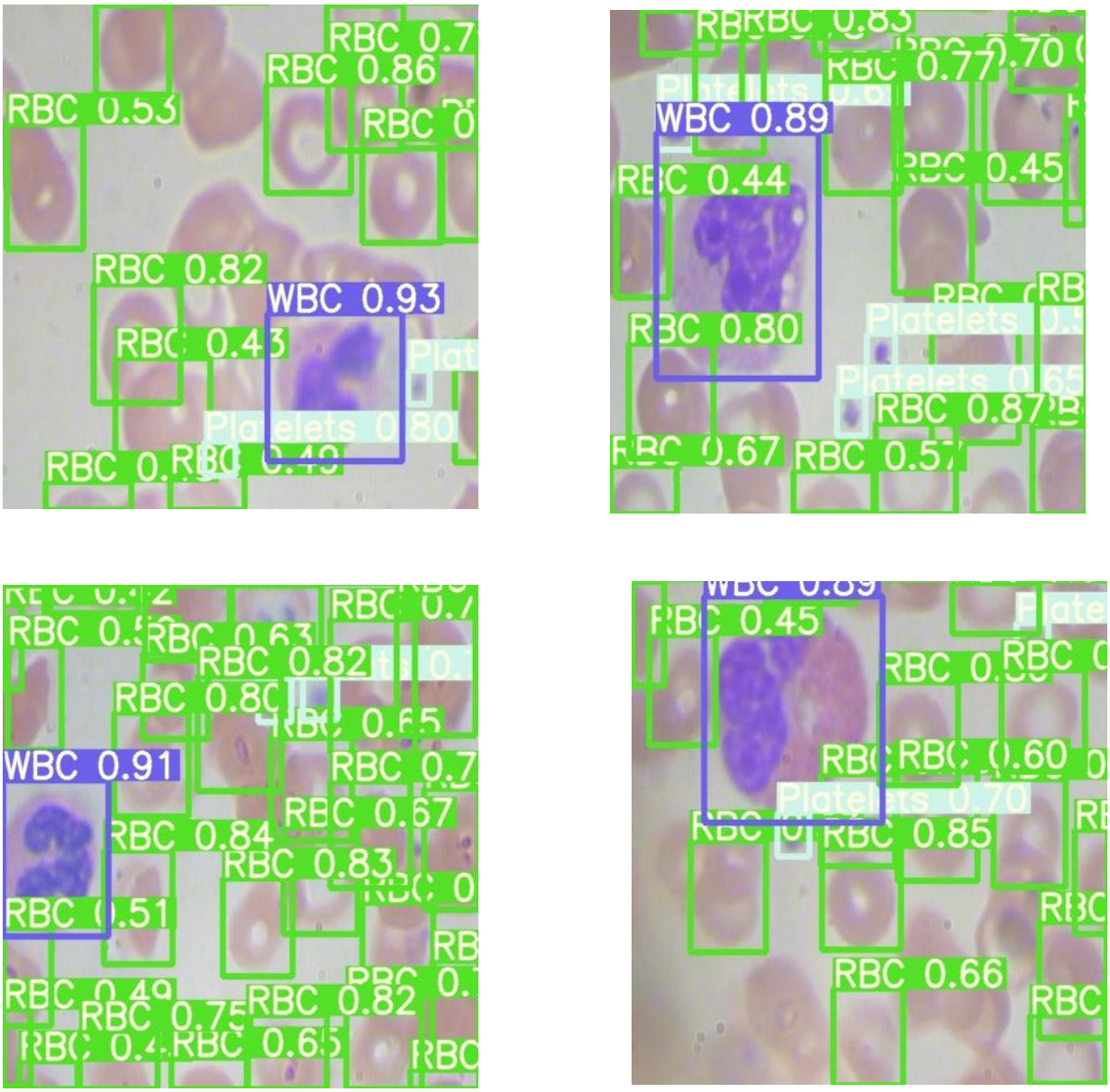


Figure. 4 Detecting accuracy result

©Daffodil International University

5.1 Findings:

A machine learning strategy based on the YOLO method is provided in this Letter to automatically recognize and count blood cells from a smear image. The method used KNN and IOU based methods to eliminate multiple counting of the same object, which improved accuracy. On publicly available datasets, we test our proposed technique. For the test dataset, it was discovered that our technique correctly identified RBCs, WBCs, and Platelets. It can be observed that our technique can reliably count even those of the cells in the dataset that aren't labeled. In the YOLO backend, various neural network models have been tested, and it has been discovered that different models can deliver the best accuracy on different cells. Despite the fact that multiple models with varied depths have been explored, the method has been found to be quite fast for counting and tagging smear images. The proposed method was also tested on a separate dataset of smear photos, and it performed admirably. With the proposed method's accuracy and detection performance, it can be said that it has the potential to simplify the manual blood cell identification and counting process.

5.2 Recommendation for Future works:

In this research Almost all of the experiments in this study were carried out with the help of we try apply more algorithms to find accuracy with others algorithms. And compare them which give the best accuracy

REFERENCES

[1] Cruz, D., Jennifer, C., Castor, L.C., Mendoza, C.M.T., Jay, B.A., Jane, L.S.C. and Brian, P.T.B., 2017, December. Determination of blood components (WBCs, RBCs, and Platelets) count in microscopic images using image processing and analysis. In *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)* (pp. 1-7). IEEE.

[2] Habibzadeh, M., Krzyżak, A. and Fevens, T., 2013, June. White blood cell differential counts using convolutional neural networks for low resolution images. In *International Conference on Artificial Intelligence and Soft Computing* (pp. 263-274). Springer, Berlin, Heidelberg.

[3] American Society of Haematology: https://www.hematology.org/education/patients/blood-basics

[4] Acharya, V. and Kumar, P., 2018. Identification and red blood cell automated counting from blood smear images using computer-aided system. *Medical & biological engineering & computing*, *56*(3), pp.483-489.

[5] Acharjee, S., Chakrabartty, S., Alam, M.I., Dey, N., Santhi, V. and Ashour, A.S., 2016, March. A semiautomated approach using GUI for the detection of red blood cells. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (pp. 525-529). IEEE.

[6] Alam, M.M. and Islam, M.T., 2019. Machine learning approach of automatic identification and counting of blood cells. *Healthcare technology letters*, *6*(4), pp.103-108.

[7] Islam, M.T., Aowal, M.A., Minhaz, A.T. and Ashraf, K., 2017. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. *arXiv preprint arXiv:1705.09850*.

[8] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. and Kim, R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, *316*(22), pp.2402-2410.

[9] Avendi, M.R., Kheradvar, A. and Jafarkhani, H., 2016. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical image analysis*, *30*, pp.108-119.

[10] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[11] Habibzadeh, M., Jannesari, M., Rezaei, Z., Baharvand, H. and Totonchi, M., 2018, April. Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception. In *Tenth international conference on machine vision (ICMV 2017)* (Vol. 10696, p. 1069612). International Society for Optics and Photonics.

[12] Xu, M., Papageorgiou, D.P., Abidi, S.Z., Dao, M., Zhao, H. and Karniadakis, G.E., 2017. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLoS computational biology*, *13*(10), p.e1005746.

[13] Yan, B., Fan, P., Lei, X., Liu, Z. and Yang, F., 2021. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sensing*, *13*(9), p.1619.

[14] Roboflow. How to Train YOLOv5 on Custom Objects. 2016. Available online: https://colab.research.google.com/drive/1gDZ2 xcTOgR39tGGs-EZ6i3RTs16wmzZQ (accessed on 3 February 2021).

[15] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J., 2018. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759-8768).

[16] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I. and Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 658-666).

[17] "YOLOv5 New Version - Improvements And Evaluation", Roboflow Blog, 2021. [Online]. Available: https://blog.roboflow.com/yolov5-improvements-and-evaluation/. [Accessed: 05- Apr- 2021].