Classification of Red Blood Cell Morphology Using Image Processing and Support Vector Machine

Paul Daniel C. Divina
School of Electrical, Electronics, and Computer Engineering
Mapúa University
1002 Muralla St., Intramuros, Manila
pdcd@mymail.mapua.edu.ph

John Philip T. Felices
School of Electrical, Electronics, and Computer Engineering
Mapúa University
1002 Muralla St., Intramuros, Manila
jptfelices@mymail.mapua.edu.ph

Carlos C. Hortinela IV
School of Electrical, Electronics, and Computer Engineering
Mapúa University
1002 Muralla St., Intramuros, Manila
cchortinela@mapua.edu.ph

Janette C. Fausto
School of Electrical, Electronics, and Computer Engineering
Mapúa University
1002 Muralla St., Intramuros, Manila
jcfausto@mapua.edu.ph

Flordeliza L. Valiente
School of Electrical, Electronics, and Computer Engineering
Mapúa University
1002 Muralla St., Intramuros, Manila
flvaliente@mapua.edu.ph

Jessie R. Balbin
School of Electrical, Electronics, and Computer Engineering
Mapúa University
1002 Muralla St., Intramuros, Manila
jjrbalbin@mapua.edu.ph

ABSTRACT

Blood serves as an indicator of health, a complete blood count (CBC) provides the clinician a view of the blood components. Diagnosis of the shape of RBC contributes information about relevant pathological diseases and condition. Red blood cells vary from the size of the cell, variations of shape, and presence of central pallor. With these observations several diseases and conditions showed correlations with the characteristic morphologic variations of red blood cells. The conventional use of peripheral blood smears remains laborious, time-consuming procedure and the lack of expertise of the microscopist is a factor to inaccurate results. The advancement of technology provided the medical field the benefits of an automated recognition and numerous studies are applying different methods in classifying RBCs. Every study differs, in what algorithm to use and what RBC is to be identified. Most of all studies were able to provide a satisfiable output and so continuous studies and development are being applied. This paper aims to develop a system that will correlate associated anemia conditions once the red blood cell was identified having an abnormality with its variations in shape or size. The proposed system was able to develop a reliable system to identify 7 different type of red blood cells normal, echinocytes, elliptocytes, dacrocytes, spherocytes, target cells, stomatocytes, and an unknown each cell achieved an accuracy of 98.33%,100%, 98.33%, 97.5%, 100%, 100%, 99.17%, and 95% respectively.

ICBET 2020, September 15–18, 2020, Tokyo, Japan

ACM ISBN 978-1-4503-7724-9/20/03…$15.00
DOI: https://doi.org/10.1145/3397391.3397407

CCS Concepts

• **Computing methodologies → Feature selection;**
• **Computing methodologies → Image processing**

Keywords

Red blood cells; Support vector machine; Conventional microscopy; Image processing; Red blood cell morphology

1. INTRODUCTION

In medical area, diagnosis of the shape of RBC contributes information about relevant pathological diseases and condition. [1]. Further, the study of blood through blood cell segmentation and blood cell identification are vital task as a health indicator [2]. A complete blood count (CBC) provides the clinician a view of the blood components which includes red blood cell (RBC), white blood cell (WBC), and platelet counts, which is the most commonly requested tests in a diagnostic laboratory [3]. In the field of health sciences, ever since the microscope was introduced, microscopic study of blood components has brought observations in the characteristics of red blood cells, from the size of the cell, variations of shape, and presence of central pallor. With these observations several diseases and conditions showed correlations with the characteristic morphologic variations of red blood cells [4]. The conventional use of peripheral blood smears remains laborious and time-consuming procedure that takes several minutes from slide preparation to microscopic analysis and the lack of expertise of the microscopist is a factor to incomplete or inaccurate results [5]. In recent years, the advancement of technology provides benefits in the medical field as automated recognition becoming more accessible and numerous studies are applying different methods in classifying RBCs. A study classified different RBCs and a sickle cell and applied a multi-layer perceptron model, the model classified WBCs and overlapped cells with 81% accuracy [6]. Another study uses SVM to classify RBC as normal or abnormal and achieved a good accuracy rates of sensitivity, specificity and Kappa to be 100%, 0.998% and 0.9944 respectively [7]. This study aims to identify red blood cells using image processing techniques and classify the cell with the help of support vector machine. This study will help correlate diseases related to the red blood cells more efficiently

since it will lessen human intervention of analyzing the specimen. It will also be beneficial to doctors, pathologist, hematologist, medical technicians, other medical related professions and especially to the patients. The device will be able to provide aid in analyzing red blood cells to help identify blood abnormality conditions in a much shorter period. In that way, preventive measures and proper medication can be immediately provided to the patient that was diagnosed which is necessary to avoid worsening the condition.

2. METHODOLOGY

The researchers have selected the problem about the conventional microscopy analysis of blood smear. The process is laborious, time-consuming, and with lack of expertise it will lead to inaccuracy. These problems can be solved with the advancement of technology becoming more accessible with an automated system of recognizing different red blood cells. Researchers obtained a comprehensive understanding of the area of study by reviewing several related literatures that could help in developing a solution to the problem. After gathering information, the researchers come up with a solution to create the conceptual framework which involves using image processing techniques such as Sobel operator, Watershed algorithm and Support vector machine. The development of the hardware and software was then designed. For the hardware development, the main components to be used are listed including the use of raspberry pi and raspberry pi camera module. The software development will be done using Python OpenCV. After developing both hardware and software, implementation will be done to create the system. Then, data gathering will follow which involves testing the system. The system will be trained with different samples. Conclusion has been made based on the data and results produced by the system.

2.1 Conceptual Framework

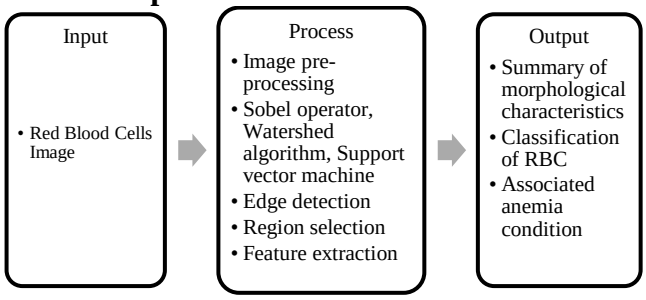


Figure 1. Conceptual framework.

The conceptual framework is shown in Figure 1. An image of red blood cell will be used as input. The inputted image will then proceed to multiple processes including image pre-processing, edge detection, image segmentation, and feature extraction. Using the algorithms Sobel operator and Watershed algorithm. At image pre-processing, the captured image will be converted into grayscale image and de-noising of input image [8]. Edge detection is applied on binary image of blood cells and recognizes gaps in images. These detected edges of cells are overlapped with binary mask of cells, and watershed transform is used for separation of overlapping cells [9]. The region selection and feature extraction will compute for the parameters area and perimeter. After all these processes were executed, support vector machine will be used as the classifier to determine the type of abnormal red blood cells present on the image. The output that will be produced are the summary of the blood morphological characteristics, classification of the red blood cells, and the associated anemia conditions. The characteristics that will be included in the summary are area, perimeter, diameter, shape geometric factor (SGF), deviation value, central pallor, and target flag. Based on the determined abnormal RBC, the system will show the possible associated types of anemia and other blood conditions. Some of these are iron-deficiency anemia, thalassemia, hereditary spherocytosis, and myelophthisic anemia.

2.2 Hardware Design

The hardware block diagram is presented on Figure 2 while in Figure 3 it shows the actual set-up of the device. A microscope is needed to view the microscopic morphology of RBCs. The raspberry pi camera is needed to capture the image from the sample, and it will be sent to the raspberry pi. A lighting will be used to provide enough lighting for the setup which is already provided on the microscope. The raspberry pi will be the main component of the system which will be the processing unit to run the program. The LCD monitor is used as a display for the user. The mouse is needed to control the raspberry pi. The power supply is applied to provide power to all the hardware components.

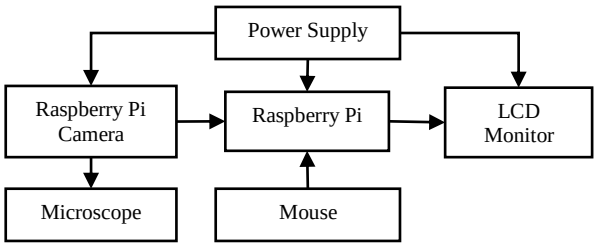


Figure 2. Hardware block diagram.

Figure 3. Hardware device set-up.

2.3 Dataset

RBCs' are extracted based on its geometrical properties [10]. Red blood cells abnormalities may be differed based on variation in size (anisocytosis), variation in shape (poikilocytosis), and variation in color [11]. Figure 4 show the dataset used for this study.

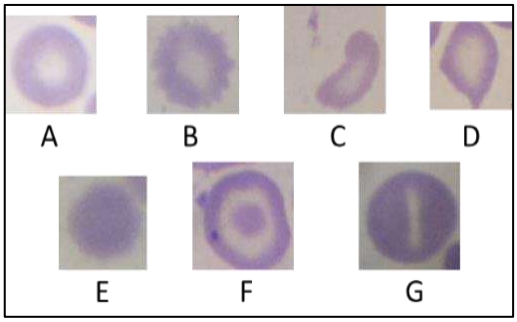


Figure 4. Dataset : a) Normal; b) Echinocytes; c) Elliptocytes; d) Dacrocytes; e) Spherocytes; f) Target Cells; g) Stomatocytes.

2.4 RBC Parameters

Different parameters are used in classifying red blood cells. The extracted parameters are used to calculate the other parameters to classify each RBCs. This is shown on Table 1.

Table 1. List of RBC features [12]

Feature	Description
Area (A)	Total number of pixels in the cell boundary
Perimeter (P)	Total number of perimeter pixels
Diameter (D)	Ratio between A and P (D = A / (4*P))
Shape Geometric Factor (SGF)	Proportion of peripheral oval's diameter
	SGF = Large diameter/Small diameter
Deviation Value (DV)	Ratio between SGF and A (DV = SGF / A)
Central Pallor (CP)	1 Central pallor is present
	0: No central pallor present
Target Flag (TF)	1: Dark pixels in the middle of central pallor
	0: No target flag present

2.5 Classification of Red Blood Cells

Table 2 shows the different types red blood cells are classified using the parameters of the red blood cells as the data set of the system.

Table 2. Red blood cells classification according to parameters

Name	D	SGF	CP	TF
Normal RBC	>6.0 & <8.0	<1.2	1	-
Echinocytes	>3.0 & <8.0	<1.2	1	-
Elliptocytes (Elongated)	-	>1.8	1	-
Dacrocytes	-	<1.8 &>1.2	1	-
Spherocytes	>3.0 & <8.0	<1.2	0	-
Target Cells	>3.0 & <8.0	<1.2	1	1
Stomatocytes	>3.0 & <8.0	<1.2	1	-

3. SOFTWARE DEVELOPMENT

Figure 5 is the system flowchart. It will initialize first some imports and set the camera resolution then the system can now acquire the image. The acquired image will undergo the processing of Sobel Edge Detection and Watershed Segmentation. Then features will be extracted from that and the RBC classification with the use of Support Vector Machine will classify the identified RBCs in the image.

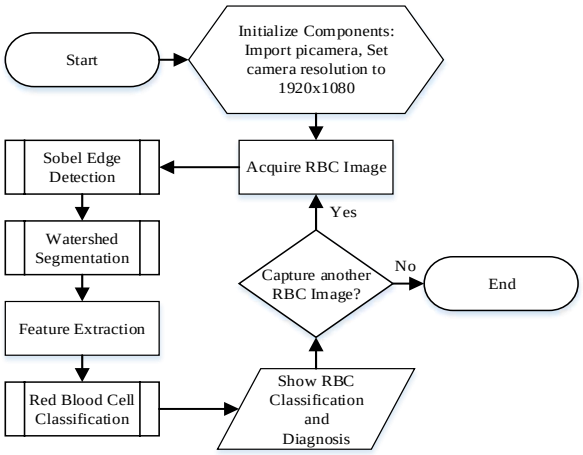


Figure 5. System flowchart.

3.1 Algorithm Design

Figure 6 show the flowchart for applying Sobel edge detection that will convert the image into grayscale to easily locate edges. Then filtering out noise or unwanted regions of the image. After that, highlight Edges for enhancement to emphasize the boundary of cell. Lastly, finding the minimum and maximum gradient intensity in which the change direction of the increase of light and dark. The next process is applying watershed segmentation, it will mark foreground objects in the image which are overlapping objects. Then background objects which are not used will be mark. Lastly, modifying of segmentation to mark background and foreground locations.

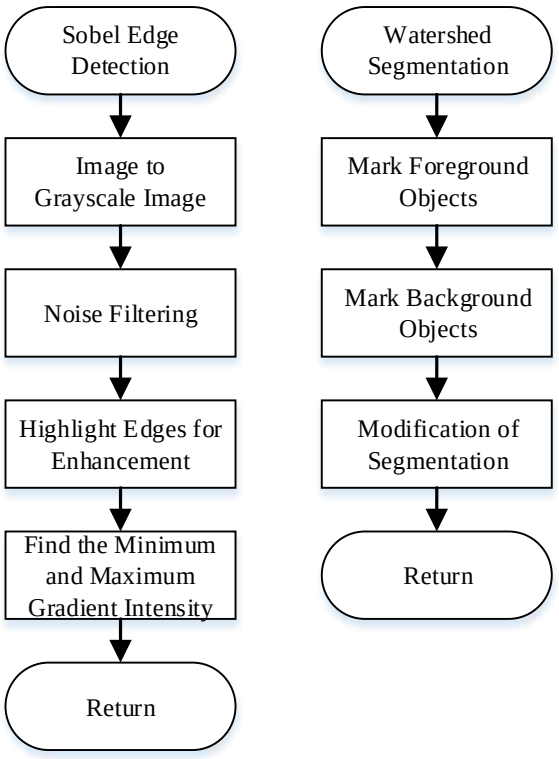


Figure 6. Flowchart of sobel edge detection and watershed segmentation algorithm.

3.2 Graphical User Interface (GUI)

The user interface of the system is presented on Figure 7. This is where the user can interact with the system. The "Capture" button is used to capture a new image. The "Select Image" button is used to view and test previously captured image. The "Change View" button is used to change between the captured blood smear image and single cell view. By going to the single cell view, you can check all the cells that are detected by the system. You can also specifically look for a specific cell and check it using the "Select Single" button.

Figure 7. GUI.

3.3 RBC Classification

The system was developed using Support Vector Machine, it will support the application to identify the variations of red blood cells. Support vector machine (SVM) is a supervised machine learning and a known dataset is used to build a model, the datasets input is trained are known and responses as the output. The system then learns from inputs and outputs from the known data set [13]. SVM has an algorithm called hyperplane enabling it to analyze patterns. Hyperplane separates different classification data for different class. [14].

The RBC will be classified according to their SGF, CP, and TF. If an image is abnormal and SGF is less than 1.2 but not less than 1.8 it is elliptocytes if not dacrocytes. If SGF is not less than 1.2 and CP does not exist, it is Spherocytes. If CP and TF exist, then it is Target Flag If it has thorny projections, it is Echinocytes and if not Stomatocytes. Figure 8 shows this process and Figure 9 shows the images acquired in the process.

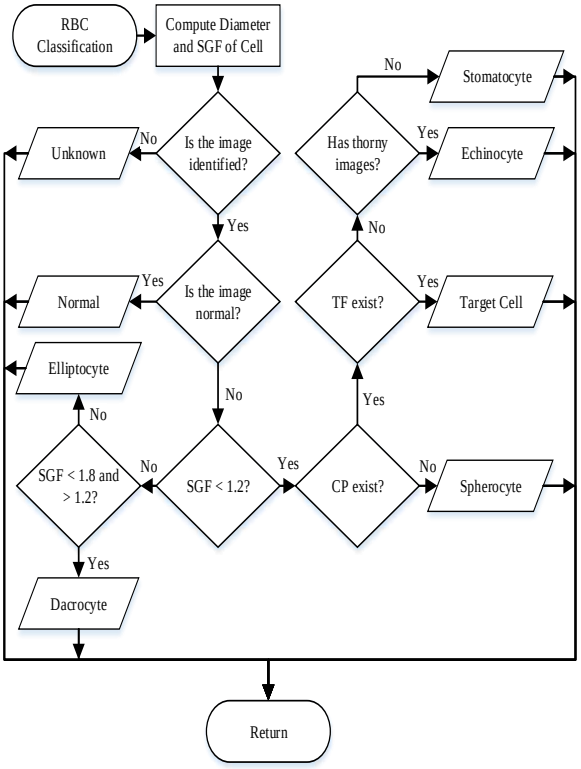


Figure 8. RBC classification.

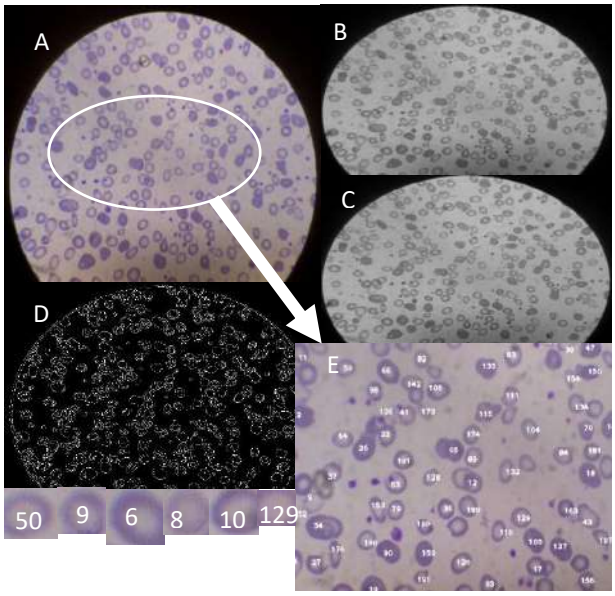


Figure 9. Acquired blood smear images and image processing algorithm utilized by the software. Shown here are the original image (A), Grayscale image (B), watershed segmentation (C), edge detection (D) and outputted cells identified by the software (E).

4. RESULTS AND DISCUSSION

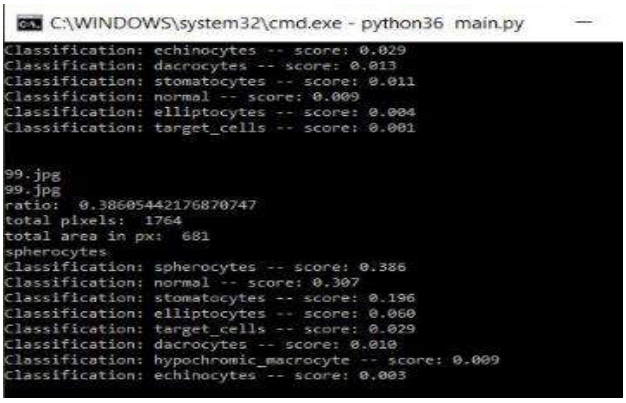
4.1 Red Blood Cell Testing Table

Table 3 shows RBC testing table each class has been tested for 15 trials. Presented on the table are some of the acquired values of morphological properties used as parameters by the system and the predicted class.

Table 3. Red blood cells testing table

Diagnose	D	SGF	CP	TF	Predicted
Normal	6.27	1.08	0	1	Normocyte
Normal	6.27	1.08	0	1	Stomatocytes
Normal	5.28	1.08	0	1	Normocyte
Normal	4.95	1.07	0	0	Spherocytes
Normal	6.93	1.08	0	1	Normocyte
Echinocytes	6.6	1.09	1	0	Echinocytes
Echinocytes	5.92	1.13	1	0	Echinocytes
Elliptocytes	3.3	2.44	1	0	Elliptocytes
Elliptocytes	5.94	2.4	1	0	Elliptocytes
Elliptocytes	0	0	1	0	Unknown
Elliptocytes	0	0	1	0	Unknown
Elliptocytes	4.29	2.42	1	0	Elliptocytes
Dacrocytes	0	0	0	0	Unknown
Dacrocytes	5.28	1.23	1	0	Dacrocytes
Dacrocytes	5.61	1.23	1	0	Unknown
Dacrocytes	4.62	1.23	1	0	Dacrocytes
Dacrocytes	5.28	1.24	1	0	Dacrocytes
Dacrocytes	7.92	1.23	1	0	Unknown
Dacrocytes	5.94	1.22	1	0	Dacrocytes
Spherocytes	5.28	1.1	1	0	Spherocytes
Target Cell	4.95	1.13	1	1	Target Cell
Target Cell	7.26	1.09	1	1	Target Cell
Stomatocytes	6.93	1.08	0	1	Stomatocytes
Stomatocytes	6.93	1.07	0	1	Normocyte
Stomatocytes	6.27	1.08	0	1	Stomatocytes
Unknown	0	0	0	0	Unknown
Unknown	0	0	0	0	Unknown

In Figure 7 in the GUI it shows the result including the parameters and the classification and the associated conditions.



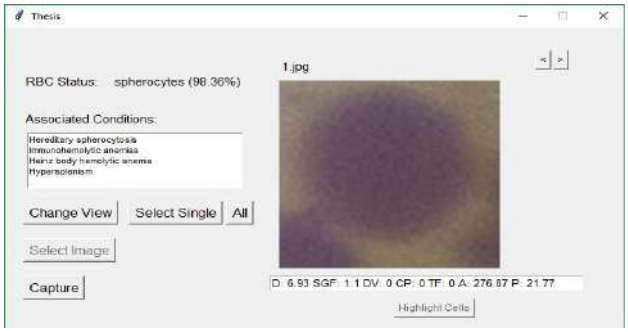


Figure 10. Graphical user interface and results of each identified cell.

Table 4. Confusion matrix for RBC classification

n = 120		Predicted RBC Class								
		Normal	Echinocytes	Elliptocytes	Dacrocytes	Spherocytes	Target Cells	Stomatocytes	Unknown	Total
Actual RBC Class	Normal	14	0	0	0	0	0	1	0	15
	Echinocytes	0	15	0	0	0	0	0	0	15
	Elliptocytes	0	0	13	0	0	0	0	2	15
	Dacrocytes	0	0	0	12	0	0	0	3	15
	Spherocytes	0	0	0	0	15	0	0	0	15
	Target Cells	0	0	0	0	0	15	0	1	15
	Stomatocytes	1	0	0	0	0	0	14	0	15
	Unknown	0	0	0	1	0	0	0	14	15
	Total	15	15	13	13	15	15	15	20	120

The table 4shows the result of the testing trials for each type of red blood cells classification included in the system. Each type has been tested for 15 trials. The result for the testing is summarized on the table. The testing is evaluated using the confusion matrix. This way of data analysis is used to create a relation and assessment on the classifier outputs of multiple classes. The confusion matrix categorizes the outputs for each testing into all the different possible results. Classifying of red blood cell can only be done one class at a time. True positive (TP) output is the total number of classifications when the system correctly identified the RBC class that is being tested. True negative (TN) output is the total number of classifications of other classes not classified as the RBC being tested. False positive (FP) output is the total number of classifications when the system incorrectly identified that the RBC belongs to the class being tested. False negative (FN) output is the total number of classifications when the system incorrectly classified that the RBC belongs to other class, but it is for the class being tested. A 2x2 confusion matrix is used to acquire the accuracy of each class. This confusion matrix will compare the result of a single class to all other classes. The accuracy for each classification can be evaluated using the formula in equation (1):

$$Accuracy = \frac{\Sigma\, TP + \Sigma\, TN}{\Sigma\, TP + \Sigma\, TN + \Sigma\, FP + \Sigma\, FN}\ x100\% \quad (1)$$

The result shows that the system has high accuracy in determining the class of red blood cell acquiring an accuracy of 98.33% for normal RBC, 100% for echinocytes, 98.33% for elliptocytes, 97.5% for dacrocytes, 100% for spherocytes, 100% for Target cells, 99.17% for stomatocytes and 95% for unknown class. Most of the errors produced are from the unknown class getting the lowest accuracy. This is mainly caused by using low quality images for testing which the system considers as unknown cells even if it is belongs to a certain class.

5. CONCLUSION

With the study, it demonstrated a great performance in the recognition of different red blood cells, according to its size and shape. Based on our testing, the software can be an aid for medical personnel as for support for there clinical reviews.

The study still needs the expertise of microscopy in locating the red blood cells. Thus, the system will only identify a red blood cell once it was captured. As the system was develop using machine learning, it can further develop to increase its accuracy. In the study few trials were only made, and it was always with the supervision of the doctor. Increasing the trials to test the accuracy of each cell can be done, gathering more datasets, identifying the color variation of red blood cell, and adding more parameters in identifying the cells will not only improve the accuracy but also the image processing and image segmentation of red blood cell.

6. ACKNOWLEDGMENTS

Our thanks to the chairman of Philippine General Hospital Department of Laboratories Dr. Nelson T. Geraldino for allowing us to use their equipment and samples in their laboratories and for assigning Dr. Erick Yturralde in assisting, for giving us some advice, and for guiding us through our study.

7. REFERENCES

[1] Tomari, R., Nurshzwani, W., & Ngadengon, R. (2015). Red blood cell counting analysis by considering an overlapping constraint. Asian Research Publishing Network, 10(3).

[2] Kolhatkar, D., & Wankhade, N. (2016). Detection and counting of blood cells using image segmentation: A review. 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave). doi:10.1109/startup.2016.7583931

[3] Ward, P. C. (2000). The CBC at the turn of the millennium: an overview. Clinical chemistry, 46(8), 1215-1220.

[4] De Vet, H. C., Koudstaal, J., Kwee, W. S., Willebrand, D., & Arends, J. W. (1995). Efforts to improve interobserver agreement in histopathological grading. Journal of clinical epidemiology, 48(7), 869-873.

[5] Ford, J. (2013). Red blood cell morphology. International journal of laboratory hematology, 35(3), 351-357.

[6] Ramin Soltanzadeh, Hossein Rabbani "Classification of three types of red blood cells in peripheral blood smear based on morphology" 978-1-4244-5900-1/10/$26.00 ©2010 IEEE.

[7] Akrimi, J. A., Suliman, A., George, L. E., & Ahmad, A. R. (2014). Classification red blood cells using support vector machine. Proceedings of the 6th International Conference on Information Technology and Multimedia. doi:10.1109/icimu.2014.7066642

[8] Man Yan, Jianyong Cai, Jiexing Gao, Lili Luo, "K-Means cluster algorithm based on color image enhancement for cell segmentation," in Proc. on Biomedical Engineering and Informatics (BMEI), IEEE, pp 295-299, 2012.

[9] Gonzalez, Rafael C., and Richard E. Woods, "Digital image processing." Prentice Hall (2002): 299-300.

[10] Tomari, R., Zakaria, W. N. W., Jamil, M. M. A., Nor, F. M., & Fuad, N. F. N. (2014). Computer Aided System for Red Blood Cell Classification in Blood Smear Image. Procedia Computer Science, 42, 206–213. doi: 10.1016/j.procs.2014.11.053

[11] Keohane, Smith, and Walenga (2016). "Haematology Clinical Principles and Applications Fifth Edition"

[12] Acharya, V., & Kumar, P. (2017). Identification and red blood cell classification using computer aided system to diagnose blood disorders. 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). doi:10.1109/icacci.2017.8126155

[13] Caya, M.V.C., Magwili, G.V., Agulto, D.L., Laranang, R.J., Palomo, L.K.G. Supervised machine learning-based fall detection (2019) 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, HNICEM 2018, art. no.8666437.

[14] Anna Monica M. De Los Reyes, Reyes, A. C., Torres, J. L., Padilla, D. A., & Villaverde, J. (2016). Detection of Aedes Aegypti mosquito by digital image processing techniques and support vector machine. 2016