AUTONOMOUS INTERPRETING PERIPHERAL BLOOD FILM BASED ON DEEP LEARNING ALGORITHM

NUR ANISAH BINTI SALEHUDDIN

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer and Microelectronic Systems)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2020

DEDICATION

This project report is dedicated to my mother, who taught me to not give up on my master programme and for giving me ideas and motivation. My father who always provide me the support even though he is not always together with us. Also, not to forget my Creator, the source and owner of my knowledge and inspiration and without his blessing I might be in difficulties to complete this project and to anyone who finds this project is helpful in any ways

ACKNOWLEDGEMENT

First and foremost, I would like to extend my gratitude towards my one and only God for his blessing and merciful help throughout my life and especially during the period of completing this Master Project. Next, to my family, my mother Maziana Osman, my dad, Salehuddin, my brother and sister, Tiha, Dhiya, Naja and Ilham in giving me countless encouraging words and keep praying for me. Also, to my Supervisor, Dr Mohd Afzan bin Othman for keep giving me insight and brilliants ideas throughout completing this project report. Thank you for giving your valuable time in giving good advice and supervise me even though mainly the communication was done through email and messenger. Without the help, I might not even get to present this project report as same as these. Also, to Dr Siti Fatimah for extending her help with the experience as a haematologist in verifying the blood count for this project. Not to forget, I am also indebted to Intel Microelectronics for funding my master programme and for providing me chance to further my studies. I would like to thanks UTM Library for giving me access in relevant articles and literatures that are useful for the research. Last but not least I want to give thanks and dedicate my heartful gratitude towards my fellow postgraduate classmate, Faizah, Syafiq and Adli, Nadhirah and Loy for your moral supports when I am at my lowest and your big heart in sharing the knowledge with me.

ABSTRACT

The peripheral blood film (PBF) is a laboratory work-up that involves cytology of peripheral blood cells smeared on a slide. As basic as it is, PBF is invaluable in the characterization of various clinical diseases as the PBF is an informative haematological tool at the clinician's disposal in screening, diagnosis and monitoring of disease progression and therapeutic response. Common clinical indication for PBF includes unexplained cytopenia, anaemia, unexplained jaundice, chronic myeloid leukaemia, suspected organ failure such as renal disease, liver failure, lymphoma and chronic lymphocytic leukaemia. PBF can only be interpreted under the microscope. A quick assessment of a PBF can be made within 3 minutes by a skilled laboratory physician but an abnormal film would require a longer time for wider view and differential cell counts. In addition, with the increasing amount of PBF screening (up to hundreds) samples requested per day, it is impossible for the laboratory physician to finish up the PBF screening within the given time frame. Besides, this conventional method tends to give inconsistent outcome as well as poor accuracy due to the significant level of inter-observer variation in grading. In Malaysia particularly, the PBF screening only available in selected General Hospital who has Hematopathology unit. Thus, all PBF samples from Klinik Kesihatan and District Hospital will be sent out to this hospital. The process itself is time consuming and tedious. Therefore, this project is aimed for the PBF to be analysed by a system that could differentiate the component on PBF which are, red blood cell (RBC), white blood cell (WBC) and platelets quantitively. Faster R-CNN algorithm for object detection is implemented as the deep learning framework for training, validating and testing the PBF images. The framework is built by integrating the Keras object detection package on top of backbone, Tensorflow library with Python as the programming language.

ABSTRAK

Filem darah periferal (FDP) adalah kerja makmal yang melibatkan sitologi selsel darah perifer yang dioleskan pada slaid. FDP adalah sangat bernilai didalam bidang klinikal dimana ianya berfungsi sebagai alat hematologi bermaklumat dalam pemeriksaan, diagnosis dan pemantauan perkembangan penyakit dan tindak balas terapeutik. Petunjuk klinikal yang biasa untuk FDP termasuk sitopenia yang tidak dapat dijelaskan, anemia, penyakit kuning yang tidak dapat dijelaskan, leukemia myeloid kronik, kegagalan organ yang disyaki seperti penyakit buah pinggang, kegagalan hati, limfoma dan leukemia limfositik kronik. FDP hanya boleh ditafsirkan di bawah mikroskop sahaja. Penilaian FDP yang pantas boleh dibuat dalam masa 3 minit oleh seorang doktor makmal yang mahir, tetapi filem yang tidak normal memerlukan masa yang lebih lama dengan melihat gambaran besar dan lebih luas untuk bilangan sel yang tepat. Di samping itu, dengan peningkatan jumlah pemeriksaan FDP sehingga beratus-ratus sampel yang diminta setiap hari, adalah mustahil bagi doctor makmal untuk menamatkan pemeriksaan dalam tempoh masa yang ditetapkan. Disamping itu, kaedah konvensional ini cenderung untuk memberikan hasil yang tidak konsisten serta ketepatan yang tidak baik kerana perbezaannya dalam pengredan. Di Malaysia terutamanya, pemeriksaan FDP hanya terdapat di Hospital Umum terpilih yang mempunyai unit Hematopatologi. Oleh itu, semua sampel FDP dari Klinik Kesihatan dan Hospital Daerah akan dihantar ke hospital ini. Proses itu sendiri memakan masa dan rumit. Juster, projek ini bertujuan untuk menganalisis FDP dengan sistem yang boleh membezakan komponen FDP iaitu sel darah merah (SDM), sel darah putih (SDP) dan platelet secara kuantitatif. Algoritma untuk pengesanan objek iaitu, Faster R-CNN dipilih sebagai rangka pembelajaran mendalam untuk latihan, pengesahan dan ujian gambar-gambar FDP. Rangka ini diintegrasi bersama pakej pengesanan objek, Keras dan tulang belakang system, Tensorflow dengan menggunakan bahasa pengaturcaraan Python.

.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PBF	-	Peripheral Blood Film
RBC	-	Red Blood
WBC	-	White Blood Cell
CBC	-	Complete Blood Count
RCD	-	Randomized Circle Detection
RoI	-	Region of Interest
BCCD	-	Blood Cell Count and Detection
CSV	-	Comma Separated Values
CNN	-	Convolutional Neural Network
R-CNN	-	Region-based CNN
RPN	-	Region Proposal Network
NMS		Non-maximum suppression
IoU		Intersection-over-Union

LIST OF SYMBOLS

n	Spatial windows distance

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Problem Background

Blood is an important fluid that circulates in human body specifically known as peripheral blood that is vital for humans' oxygen supply, immunity, nutrition and waste secretion [1]. A normal human will have about 4.5 to 5.5 litres of blood inside their body [2]. Thousand and billions of cells are composed inside blood that deliver different function to the body and the major cells are red blood cell, white blood cell, and platelets (Figure 1.1). These cells were wrapped around together with another protein-composed fluid called plasmas.

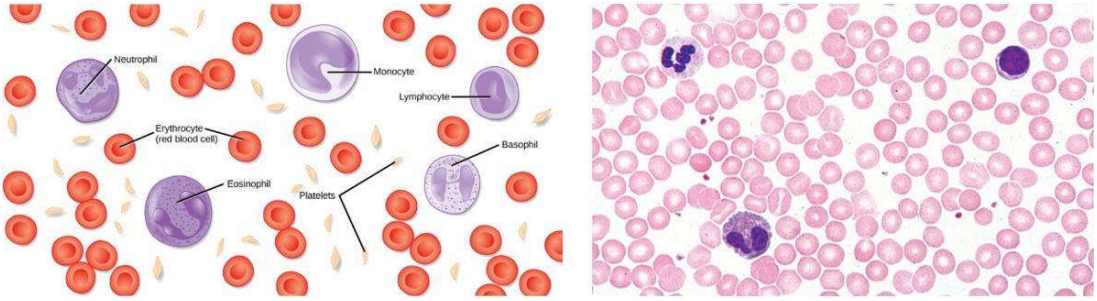

Figure 1.1 Illustration of 3 major cells in the peripheral blood (left) and Peripheral Blood Film (right).

The diagnosis of this peripheral blood film helps in determine the health condition of a person. This is done by having the peripheral blood smeared on a slide with utmost care following the SOP as to have the specimen in good condition when diagnosed later. The specimen will be placed under microscope to gets its image as peripheral blood film (Figure 1.1). Specialist in charge of making the deduction throughout the diagnosis is known as haematologist. They will gather the useful information on the blood cell morphology from the PBF for example, the cell counts, volume or its variation in shape and size [3]

1.2 Blood Cell Count

A simple blood test is done to acquire the complete blood count (CBC) to monitor the possibility of the patient having a disease. According to Laura [4], the RBC cell count for a normal adult is different between female and male. Normally, the red blood cell will have a higher count than white blood cell and platelets with values of 40% to 45% of its blood volume [5]. The blood count range for RBC, WBC and platelets are summarised as Table 1.1.

Table 1.1 Complete Blood Count

Blood Cell	**Reference Range**	
	Male (per mm^3)	**Female (per mm^3)**
Red Blood Cell	4.3-5.9 million	3.5-5.5 million
White Blood Cell	4500-11,000	
Platelets	150,000-400,000	

When the counts of the blood cell are not in the reference range, it can help haematologist to diagnose on the symptom to the related disease in which need to be further investigated. For example, some known disease that are related to RBC is the anaemia. Apart from RBC count, its average size, total space consumed in the blood and amount of Haemoglobin are also taken into consideration [6]. If the RBC count is below the reference range, the patient is diagnosed from having anaemia.

Meanwhile for WBC, the low counts might lead to the condition of infection, chemotherapy as to kill cancerous cells, AIDS and Lupus and high cell count points to leukaemia, a type of cancer [7]. As for the platelets, having a higher count can associates with thrombocytosis that can lead to stroke, heart attack or blood clot formed in blood vessels [8].

1.3 Problem Statement

To evaluate the PBF, a skilled laboratory physician would take up about 3 minutes of a quick assessment but for peculiar or abnormal ones would take up more 2 time for wider view and to get the correct differential cell counts [9]. The PBF diagnosis is first done by preparing a correctly made peripheral blood smear. It is then stained with a specific stainer to get the details of blood component such as its respective nuclei and cytoplasm [5].

1.4 Research Objective

1. To implement the deep learning algorithm that can identify between RBC, WBC and Platelet from PBF.

2. To have the cell count of RBC, WBC and Platelet from PBF by using deep learning model.

3. To achieve high accuracy for cells identification.

1.5 Scope of Work

1.5.1 Software and tools

This system implements Faster R-CNN deep learning model framework for object detection algorithm. The coding will be using Python programming language with Anaconda Python version 3.6. This architecture is installed with Keras package and Tensorflow Library for the object detection model. The dataset augmentation and the calculation of the blood cell will be done in Jupyter Notebook.

1.5.2 Datasets

Input dataset is collected from MIT-licensed Roboflow BCCD Dataset in two different data format for its annotation which are .xml file as Pascal VOC and .csv file as Tensorflow Object Detection [10]. The model is tested with another dataset with different image from testing and validation to get the accuracy of the prediction as well as the count for each blood cell in an image.

1.5.3 Limitation

The model will only predict three major blood component which are RBC, WBC and Platelets and would not further classify the class for each blood cell. As for the testing dataset, the image was selected only for the clear RBC shape composition.

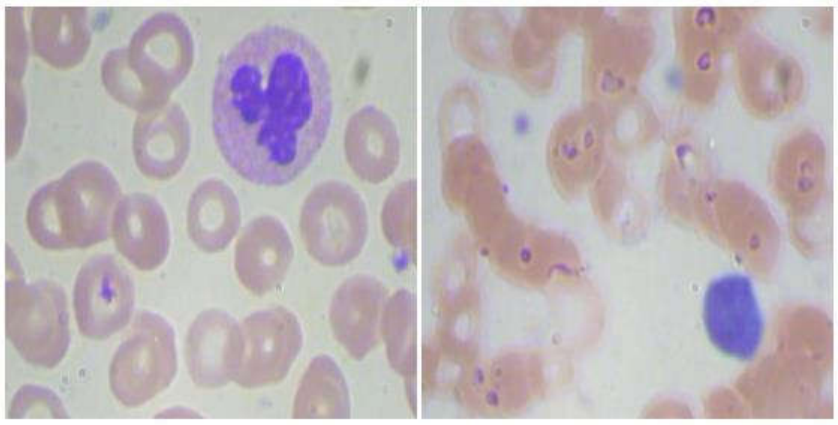

Figure 1.2 Example image used as testing dataset (left). Example image of overlapped RBC that is discarded from dataset (right).

REFERENCES

[1] J. Rodellar, S. Alférez, A. Acevedo, A. Molina, and A. Merino, "Image processing and machine learning in the morphological analysis of blood cells," *Int. J. Lab. Hematol.*, vol. 40, no. January, pp. 46–53, 2018, doi: 10.1111/ijlh.12818.

[2] R. Rettner, "The Human Body_ Anatomy, Facts & Functions _ Live Science," 2016, Accessed: Dec. 18, 2019. [Online]. Available: https://www.livescience.com/37009-human-body.html.

[3] John Meletis, "the Peripheral Blood Film in the Diagnostic Approach of Common and Rare Diseases," *5th Congr. Clin. Chem. Laboratoty Med.*, no. May, 2011.

[4] L. Dean, "Blood and the cells it contains," *Blood Groups Red Cell Antigens*, no. Md, pp. 1–6, 2005, [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK2263/.

[5] American Society if Hematology, "Hematology Glossary - Hematology.org." https://www.hematology.org/education/patients/blood-basics (accessed Jul. 03, 2020).

[6] Rice University, "Erythrocytes Anatomy and Physiology." https://opentextbc.ca/anatomyandphysiology/chapter/18-3-erythrocytes/ (accessed Jul. 03, 2020).

[7] H. J. M. de Jonge *et al.*, "Prognostic impact of white blood cell count in intermediate risk acute myeloid leukemia: Relevance of mutated NPM1 and FLT3-ITD," *Haematologica*, vol. 96, no. 9, pp. 1311–1317, Sep. 2011, doi: 10.3324/haematol.2011.040592.

[8] T. S. Njølstad, H. Engerud, H. M. J. Werner, H. B. Salvesen, and J. Trovik, "Preoperative anemia, leukocytosis and thrombocytosis identify aggressive endometrial carcinomas," *Gynecol. Oncol.*, vol. 131, no. 2, pp. 410–415, Nov. 2013, doi: 10.1016/j.ygyno.2013.08.032.

[9] A. S. Adewoyin and B. Nwogoh, "Peripheral blood film - a review.," *Ann. Ibadan Postgrad. Med.*, vol. 12, no. 2, pp. 71–9, Dec. 2014, Accessed: Dec. 17, 2019. [Online]. Available:

http://www.ncbi.nlm.nih.gov/pubmed/25960697.

[10] "Roboflow BCCD Dataset," *MIT*, 2020. https://public.roboflow.ai/object-detection/bccd (accessed Jul. 03, 2020).

[11] A. Kapoor, "Deep Learning vs. Machine Learning: A Simple Explanation," Feb. 2019. https://hackernoon.com/deep-learning-vs-machine-learning-a-simple-explanation-47405b3eef08 (accessed Jul. 03, 2020).

[12] Y. M. Alomari, S. N. H. Sheikh Abdullah, R. Zaharatul Azma, and K. Omar, "Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm," *Comput. Math. Methods Med.*, vol. 2014, pp. 1–17, 2014, doi: 10.1155/2014/979302.

[13] F. Al-Hafiz, S. Al-Megren, and H. Kurdi, "Red blood cell segmentation by thresholding and canny detector," *Procedia Comput. Sci.*, vol. 141, pp. 327–334, 2018, doi: 10.1016/j.procs.2018.10.193.

[14] A. Elen and M. K. Turan, "International Journal of Advanced and Applied Sciences A new approach for fully automated segmentation of peripheral blood smears," vol. 5, no. 1, pp. 81–93, 2018.

[15] A. Elen and M. K. Turan, "A new approach for fully automated segmentation of peripheral blood smears," *Int. J. Adv. Appl. Sci.*, vol. 5, no. 1, pp. 81–93, 2018, doi: 10.21833/ijaas.2018.01.011.

[16] A. Hamouda, A. Y. Khedr, and R. A. Ramadan, "Automated Red Blood Cell Counting," *Int. J. Comput. Sci.*, vol. 1, no. 2, pp. 13–16, 2012, [Online]. Available: http://researchpub.org/journal/ijcs/number/vol1-no2-3.pdf.

[17] E. Suryani and N. Polvonov, "Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia," vol. 2, p. 6, 2014, Accessed: Jul. 04, 2020. [Online]. Available: http://www.ijcsns.com.

[18] V. Acharya and P. Kumar, "Identification and red blood cell automated counting from blood smear images using computer-aided system," *Med. Biol. Eng. Comput.*, vol. 56, no. 3, pp. 483–489, 2018, doi: 10.1007/s11517-017-1708-9.

[19] N. F. Hassan, "Determining Numbers of Red Blood Cells in Blood Smear Images Using Blob Detection," *Irawq Academic Scientific Journals*, 2017. https://www.iasj.net/iasj?func=article&aId=125647 (accessed Jul. 05, 2020).

[20] D. Mundhra, B. Cheluvaraju, J. Rampure, and T. Rai Dastidar, "Analyzing

microscopic images of peripheral blood smear using deep learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10553 LNCS, pp. 178–185, doi: 10.1007/978-3-319-67558-9_21.

[21] M. M. Alam and M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells," *Healthc. Technol. Lett.*, vol. 6, no. 4, pp. 103–108, 2019, doi: 10.1049/htl.2018.5098.

[22] T. Tran, O. H. Kwon, K. R. Kwon, S. H. Lee, and K. W. Kang, "Blood Cell Images Segmentation using Deep Learning Semantic Segmentation," *2018 IEEE Int. Conf. Electron. Commun. Eng. ICECE 2018*, no. September, pp. 13–16, 2019, doi: 10.1109/ICECOME.2018.8644754.

[23] M. Xu, D. P. Papageorgiou, S. Z. Abidi, M. Dao, H. Zhao, and G. E. Karniadakis, "A deep convolutional neural network for classification of red blood cells in sickle cell anemia," *PLoS Comput. Biol.*, vol. 13, no. 10, pp. 1–27, 2017, doi: 10.1371/journal.pcbi.1005746.

[24] T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, and A. Phon-on, "Convolutional Neural Networks for Recognition of Lymphoblast Cell Images," vol. 2019, 2019.

[25] B. Sheng, M. Zhou, M. Hu, Q. Li, L. Sun, and Y. Wen, "A blood cell dataset for lymphoma classification using faster R-CNN," *Biotechnol. Biotechnol. Equip.*, vol. 34, no. 1, pp. 413–420, Jan. 2020, doi: 10.1080/13102818.2020.1765871.

[26] R. R. Tobias *et al.*, "Faster R-CNN Model with Momentum Optimizer for RBC and WBC Variants Classification," in *LifeTech 2020 - 2020 IEEE 2nd Global Conference on Life Sciences and Technologies*, Mar. 2020, pp. 235–239, doi: 10.1109/LifeTech48969.2020.1570619208.

[27] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11. Institute of Electrical and Electronics Engineers Inc., pp. 3212–3232, Nov. 01, 2019, doi: 10.1109/TNNLS.2018.2876865.

[28] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, "Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 382–392, 2019, doi: 10.1016/j.bbe.2019.01.005.

[29] N. Papernot, P. Mcdaniel, S. Jha, M. Fredrikson, Z. B. Celik, and A. Swami, "The limitations of deep learning in adversarial settings," in *Proceedings - 2016 IEEE European Symposium on Security and Privacy, EURO S and P 2016*, May 2016, pp. 372–387, doi: 10.1109/EuroSP.2016.36.

[30] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," 2015. Accessed: Jul. 05, 2020. [Online]. Available: https://github.com/.

[32] R. Girshick, "Fast R-CNN," 2015. Accessed: Jul. 05, 2020. [Online]. Available: https://github.com/rbgirshick/.

[33] S. Pulkit, "Introduction to Object Detection Algorithms," Oct. 2018. https://www.analyticsvidhya.com/blog/2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/ (accessed Jul. 05, 2020).

[34] Mathworks, "Getting Started with R-CNN, Fast R-CNN, and Faster R-CNN - MATLAB & Simulink." https://www.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.html#mw_25d18973-df6c-48ef-aaa9-31a4ec9e6705 (accessed Jul. 05, 2020).

[35] G. Rohith, "R-CNN, Fast R-CNN, Faster R-CNN, YOLO Object Detection Algorithms," Jul. 2018. https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e (accessed Jul. 05, 2020).

[36] Y. Ren, C. Zhu, S. Xiao, and F. Marotti De Sciarra, "Object Detection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures," 2018, doi: 10.1155/2018/3598316.

[37] D. S. F. Mohamed Kamaruzzaman, "Haematologist Specialist, Hospital Sultan Ismail, Johor Bahru Malaysia," 2020.