ENHANCING PHILADELPHIA MUTATION DETECTION USING OPTIMIZED IMAGE PRE-PROCESSING FOR CONVOLUTION NEURAL NETWORK

By

Jacinthya Stephany Ezralia Harun
11906006

BACHELOR'S DEGREE
in

ELECTRICAL ENGINEERING - BIOMEDICAL ENGINEERING CONCENTRATION
LIFE SCIENCES AND TECHNOLOGY

SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

Revision after the Thesis Defense on 17 July 2023

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

Jacinthya Stephany Ezralia Harun
Student — 31 July 2023 Date

Approved by:

Aulia Arif Iskandar, S.T., M.T.
Thesis Advisor — 31 July 2023 Date

Dr. dr. Lyana Setiawan Sp.PK(K)
Thesis Co-Advisor — 31 July 2023 Date

Dr. Hery Susanto, S.Si., M.Si.
Dean — 31 July 2023 Date

ABSTRACT

ENHANCING PHILADELPHIA MUTATION DETECTION USING OPTIMIZED IMAGE PRE-PROCESSING FOR CONVOLUTION NEURAL NETWORK

By

Jacinthya Stephany Ezralia Harun
Aulia Arif Iskandar, S.T., M.T., Advisor
Dr. dr. Lyana Setiawan Sp.PK(K), Co-Advisor

SWISS GERMAN UNIVERSITY

Leukemia is a cancer that have a high mortality rate. Identifying the marker of leukemia could help determine the type of leukemia and its treatment. Philadelphia mutation (Ph) t(9;22)(q34;q11.2) is one of the markers that commonly occurred in acute lymphocytic leukemia-ALL (20-30%) and chronic myelocytic leukemia-CML (90-95%) patients. This marker can be detected by several diagnosis tests, including Karyotyping. However, karyotyping is known to have high errors due to human factors and poor chromosomal shape and band. This research aims to aid medical professionals in detecting Ph mutation by developing an image processing algorithm using Gaussian Blur, Otsu thresholding, CLAHE, Median Blur, and Convolution Neural Network. 400 chromosome images were collected from literature studies, Dharmais Cancer Hospital, RSUPN Dr. Cipto Mangunkusumo (RSCM), and Mitra Keluarga Laboratory as the dataset. The first step was the preprocessing method followed by chromosomal classification by using fine-tuned Inception-ResNet V2 CNN. The step achieved the average accuracy, precision, recall, and F1-score of 92.88%, 84.31 %, 82.55%, and 82.77%, respectively. The final step was detecting Ph mutation and this step obtained

84.6% accuracy. It is concluded that the image processing algorithm potentially detects Ph mutation and aids the detection of Ph mutation in karyotyping.

Keywords: Leukemia, Karyotyping, Philadelphia Chromosome, Convolution Neural Network

© Copyright 2023
by Jacinthya Stephany Ezralia Harun
All rights reserved

DEDICATION

I dedicate this work to my beloved Jesus, family and friends. Furthermore, I dedicate this work to help improve the cytogenetic examination in the clinical pathology laboratory, at Dharmais Cancer Hospital.

ACKNOWLEDGEMENTS

I'd like to give praises and thanks to Jesus, my lovely teacher, and the One who always support me during my stressful time making this thesis. I am successfully finishing my study because of Him.

Secondly, I am grateful to my advisor, Mr. Aulia Arif Iskandar, S.T., M.T., and my co-advisor Dr. dr. Lyana Setiawan Sp.PK(K) from Dharmais Cancer Hospital for their ongoing mentorship throughout the making of my thesis work. I could not finish my work without them.

I would also like to thank Dharmais Cancer Hospital Staff: kak Ajeng, kak Tika, kak Maria, kak Yassin, and Bu Titing for giving me guidance, motivation as well as endless food when I was doing my research at Dharmais Hospital.

I am deeply thankful to Chandrika for her endless support and threat that triggered me to stop procrastinating. Lastly, I am grateful to have a supportive family, especially my mama and papa as my 911 and the one who never fails to calm me down.

TABLE OF CONTENTS

	Page
STATEMENT BY THE AUTHOR	2
ABSTRACT	3
DEDICATION	6
ACKNOWLEDGEMENTS	7
TABLE OF CONTENTS	8
LIST OF FIGURES	10
LIST OF TABLES	13
CHAPTER 1 - INTRODUCTION	15
1.1 Background	15
1.2 Research Problems	18
1.3 Research Objectives	18
1.4 Significance of Study	18
1.5 Research Questions	19
1.6 Hypothesis	19
CHAPTER 2 - LITERATURE REVIEW	20
2.1 Karyotyping Procedure	20
2.2 Chromosome Reading System Based on the Banding and the Region	24
2.3 Chromosomal Translocation Abnormality Reading System	26
2.4 Acute Lymphocytic Leukemia	26
2.5 Chronic Myelocytic Leukemia	28
2.6 Median Blur Filter	31
2.7 Otsu Thresholding	32
2.8 Gaussian Filter	34
2.9 Contrast Limited Adaptive Histogram Equalization (CLAHE)	35
2.10 Data Augmentation	37
2.11 Convolution Neural Network (CNN)	38
CHAPTER 3 – RESEARCH METHODS	41
3.1 Materials and Equipment	41

3.2 Study Design 49
3.3 Analytical Method 58
CHAPTER 4 – RESULTS AND DISCUSSIONS 61
4.1 Result of the Preprocessing Step 61
4.2 CNN Training and Testing Result 70
4.3 Comparison between Two Experiments 83
4.4 Results' Comparison with Other Studies 86
4.5 Limitations 91
CHAPTER 5 – CONCLUSIONS AND RECOMMENDATIONS 93
5.1 Conclusions 93
5.2 Recommendations 93
GLOSSARY 95
REFERENCES 96
APPENDIX 113
CURRICULUM VITAE 124

LIST OF FIGURES

Figure Page

Figure 1 Mitosis Process of The Cell (Gilchrist, 2023) 15

Figure 2 10 Most Frequent Cancer Cases in Indonesia in 2020 (Rizaty, 2021) 16

Figure 3 Division of Blood Stem Cell into Different type of White Blood Cells and Other Blood Cells (National Cancer Institute, n.d) 17

Figure 4 The Example of A Normal Male (46,XY) Karyotype (Thirumulu, 2011) 20

Figure 5 An ideogram of a male chromosome (Heim & Mitelman, 1986) 25

Figure 6 The chromosomal translocation t(9;22)(q34;q11) (Pérez-Caro & Sánchez-Garcia, 2007) 26

Figure 7 Median Blur Filter Output Example. Left: Original image, middle: median blur filter with radius 1 applied twice, right: median blur filter with radius 1 applied three times (GIMP, n.d) 32

Figure 8 Otsu Thresholding Examples. 34

Figure 9 Gaussian Smoothing Filter Output Examples; Original (left), kernel size :3 (middle), kernel size:10 (right) 35

Figure 10 The algorithm of Contrast Limited Adaptive Histogram Equalization (MathWorks, n.d) 37

Figure 11 Illustration of the Convolution Neural Network (Swapna, 2022) 39

Figure 12 Separation Process from Full Karyotype of an ALL Patient from Dharmais Cancer Hospital (A) to 4 Individual Chromosome Datasets (B) 43

Figure 13 Workflow of Karyotyping Chromosome 44

Figure 14 Block Diagram of the First Experiment 50

Figure 15 Block Diagram of the Second Experiment 51

Figure 16 Flow of Pre-processing Algorithm 51

Figure 17 Splitting Folder for Training and Testing 54

Figure 18 Workflow of Input Images Before Entering CNN 55

Figure 19 Splitting Folder for Training and Validation Set 56

Figure 20 Basic Architecture of Inception-Resnet-v2 56

Figure 21 CNN Multi-Class Classification Workflow 57

Figure 22 A Good Fit Example of Model Loss Curve Graph (Brownlee, 2019) 58

Figure 23 Overfitting Model Loss (left) and Underfitting Model Loss(right) Curve Graph (Brownlee, 2019) 59

Figure 24 Confusion Matrix (Sharma et al., 2022) 59

Figure 25 The 9th Chromosome (A) and 22nd Chromosome (B) after Manually Extracted 62

Figure 26 Comparison of FFT of an image before (A) and after (B) the extraction process 62

Figure 27 The histogram of Chromosome 9 63

Figure 28 Result of Gaussian Blur Filter: (A) Original image; (B) Gaussian filtered 64

Figure 29 Cropping using Otsu Thresholding 64

Figure 30 Histogram of the Input Image Before and After the Cropping Process 65

Figure 31 FFT plot of (A) Original Image; (B) after Cropping 65

Figure 32 FFT of Cropped Image (left) and CLAHE (right) 66

Figure 33 FFT of CLAHE Image (left) and Median Blurred Image (right) 67

Figure 34 Output of Padding and Resizing 68

Figure 35 Results of Augmentation Process: (A) Horizontally flipped image; (B) -90 to 90 degrees rotated image; (C) -270 to 270 degrees rotated image. 70

Figure 36 Accuracy Plot (A) and Loss Plot (B) of the First Experiment Model 71

Figure 37 Confusion Matrices after Testing the Dataset with Three Trial Models of the First Experimental Model 73

Figure 38 Confusion Matrix of the Philadelphia Classification for the 1st, 2nd and 3rd Trials 77

Figure 39 The Second Trial's Accuracy Plot (A) and Loss Plot (B) of the Second Experiment Mode 78

Figure 40 The Second Trial's Accuracy Plot (A) and Loss Plot (B) of the Second Experiment Mode 79

Figure 41 Confusion Matrices after Testing the Dataset with Three Trial Models of the Second Experimental Model 80

Figure 42 The Example of Chromosome 9 Before (A) and After (B) the Preprocessing Step 82

Figure 43 Average Comparison Between Two Experimental Model 84

Figure 44 Comparison of Two Philadelphia Classification Test 86

Figure 45 Comparison Between the Current Study and the Other Philadelphia Classification Studies 87

Figure 46 Comparison Between the Current Study and the Other Chromosomal Classification Studies 90

Figure 47 Model Summary of Inception-ResNetV2 Architecture 121

Figure 48 Training Set for CNN 123

LIST OF TABLES

Table | Page

Table 1 WHO 2018 Classification of Acute Lymphocytic Leukemia 27

Table 2 Equipment Used for Karyotyping Procedures 48

Table 3 First Pre-processing Algorithm (detailed algorithm see Appendix A) 52

Table 4 Second Pre-processing Algorithm (detailed algorithm see Appendix A) 53

Table 5 Third Pre-processing Algorithm (detailed algorithm see Appendix A) 53

Table 6 Last Classification Algorithm (detailed algorithm see Appendix C) 57

Table 7 The Summary of Images in Training and Testing Folder for Different Classes 69

Table 8 Epoch Optimization in Current Research 71

Table 9 Summary of Evaluation Metrics of the First Experiment 75

Table 10 Result of Philadelphia Classification Test for the First Experiment 76

Table 11 Summary of Philadelphia Classification Test 78

Table 12 Summary of Evaluation Metrics of the First Experiment 81

Table 13 Result of Philadelphia Classification Test for the Second Experiment 83

Table 14 Comparison of Performance Score Between Current Studies and Other Literature 88

Table 15 First Pre-processing Algorithm 113

Table 16 Padding Algorithm 114

Table 17 Resize Algorithm 115

Table 18 Splitting Algorithm 115

Table 19 Augmentation Algorithm 116

Table 20 CNN Algorithm 117

Table 21 Philadelphia Classification Algorithm 122