



**IMAGE ENHANCEMENT AND FEATURE  
EXTRACTION TECHNIQUES FOR INTELLIGENT  
THALASSEMIA SCREENING PROCEDURES**

by

**NURUL ZHAFIKHA BINTI NOOR RASHID  
(1131310627)**

A thesis submitted in fulfilment of the requirements for the degree of  
Master of Science in Biomedical Electronics Engineering

School of Mechatronic Engineering  
UNIVERSITI MALAYSIA PERLIS

2016

## ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful.

Alhamdulillah, all praises to Allah for the strengths and His blessing in completing this thesis. This thesis owes its existence to the help, support and inspiration of several people. First and foremost, I would like to express my sincere appreciation to my supervisor, Prof. Dr. Yusoff Mashor for his supervision during this research. His guidance and motivations always keep me focused on the research objectives and have been precious in the development of this thesis content.

I would also like to express my appreciation to the medical supervisor Prof. Dr. Rosline Hassan, haematologist of Hospital Universiti Sains Malaysia and all the staff for providing the constructive guidance during the research study. I would like to convey my gratitude to the Ministry of Higher Education (MOHE) for the scholarship that helps my aspiration became possible.

I would like to express my deepest gratitude to my beloved husband, Muhammad Shafiq and my precious daughter, Hana Insyirah for keep believe in me and bear with me all the times. Not forgotten to my beloved parents, Noor Rashid Hamid and Zaleha Abd Rahman, my beloved parents-in-laws, Mezan Kassim and Siti Salasiah, and the rest of my family for the prayer, love, motivation and encouragement that inspire me to strive harder in achieving the dreams.

My sincere thanks must also go to all my colleagues especially to Dr. Aimi Salihah, Dr. Rafikha Aliana and my fellow friends for all the support and guidance in terms of knowledge and motivation during this whole period. Your help and encouragement really means to me. Thank you very much.

**Nurul Zhafikha bt Noor Rashid, January 2016.**

## TABLE OF CONTENTS

|   | <b>PAGE</b>  |
|---|--------------|
| <b>THESIS DECLARATION</b>                   | <b>I</b>     |
| <b>ACKNOWLEDGEMENT</b>                      | <b>II</b>    |
| <b>TABLE OF CONTENTS</b>                    | <b>III</b>   |
| <b>LIST OF TABLES</b>                       | <b>IX</b>    |
| <b>LIST OF FIGURES</b>                      | <b>XI</b>    |
| <b>LIST OF ABBREVIATIONS</b>                | <b>XVII</b>  |
| <b>LIST OF SYMBOLS</b>                      | <b>XIX</b>   |
| <b>ABSTRAK</b>                              | <b>XXII</b>  |
| <b>ABSTRACT</b>                             | <b>XXIII</b> |
| <br>  |              |
| <b>CHAPTER 1: INTRODUCTION</b>              |              |
| 1.1 Background                              | 1            |
| 1.2 Problem Statement                       | 3            |
| 1.3 Research Objectives                     | 4            |
| 1.4 Scope of the Research                   | 5            |
| 1.5 Thesis Outlines                         | 6            |
| <br>  |              |
| <b>CHAPTER 2: LITERATURE REVIEW</b>         |              |
| 2.1 Introduction                            | 8            |
| 2.2 Thalassemia Disease                     | 8            |
| 2.2.1 Normal Red Blood Cell and Haemoglobin | 10           |

|         |  |    |
|---------|--|----|
| 2.2.2   | Types, Causes and Signs of Thalassemia                             | 11 |
| 2.2.3   | Epidemiology   | 14 |
| 2.2.4   | Morphological Characteristic of Normal and Thalassemia Blood Cells | 15 |
| 2.2.5   | Laboratory Test for Thalassemia                                    | 18 |
| 2.2.5.1 | Complete Blood Count   | 18 |
| 2.2.5.2 | Haemoglobin Electrophoresis  | 20 |
| 2.2.5.3 | High Performance Liquid Chromatography (HPLC)                      | 21 |
| 2.2.5.4 | DNA Analysis   | 21 |
| 2.2.6   | Treatment and Prevention of Thalassemia Disease                    | 22 |
| 2.3     | Digital Image Processing of Blood Images                           | 23 |
| 2.3.1   | Image Acquisition  | 24 |
| 2.3.2   | Image Enhancement for Blood Images                                 | 28 |
| 2.3.2.1 | Dark Contrast Enhancement Technique                                | 28 |
| 2.3.2.2 | Bright Contrast Enhancement Technique                              | 30 |
| 2.3.2.3 | Partial Contrast Enhancement Technique                             | 30 |
| 2.3.2.4 | Global Contrast Enhancement Technique                              | 31 |
| 2.3.2.5 | Local Contrast Enhancement Technique                               | 32 |
| 2.3.2.6 | Application of Enhancement Techniques for Blood Diseases           | 33 |
| 2.3.3   | Image Segmentation of Blood Images                                 | 34 |
| 2.3.4   | Morphological Features Extraction of Blood Cells                   | 38 |
| 2.3.5   | Feature Selection of Blood Cell Features                           | 40 |
| 2.3.5.1 | Chi-Squared Filter   | 43 |

|  |   |    |
|--|---|----|
| 2.3.5.2                                | Information Gain Filter                         | 43 |
| 2.3.5.3                                | Correlation-based Feature Selection Filter      | 44 |
| 2.3.5.4                                | Principal Component Analysis                    | 45 |
| 2.4                                    | Artificial Neural Network                       | 47 |
| 2.4.1                                  | Multilayer Perceptron Neural Network            | 48 |
| 2.4.1.1                                | Levenberg-Marquardt Algorithm                   | 50 |
| 2.4.1.2                                | Bayesian Regularization Algorithm               | 52 |
| 2.4.1.3                                | Scale Conjugate Gradient Algorithm              | 53 |
| 2.4.2                                  | Application of ANN in Medical Field             | 54 |
| 2.5                                    | Classification Procedures for Blood Cells       | 56 |
| 2.6                                    | Summary   | 59 |
| <br>                                   |   |    |
| <b>CHAPTER 3: RESEARCH METHODOLOGY</b> |   |    |
| 3.1                                    | Introduction                                    | 61 |
| 3.2                                    | Image Acquisition                               | 63 |
| 3.3                                    | Contrast Image Enhancement                      | 64 |
| 3.4                                    | Segmentation of Blood Images                    | 67 |
| 3.4.1                                  | Segmentation of Blood Cells                     | 69 |
| 3.4.1.1                                | Image Color Conversion                          | 70 |
| 3.4.1.2                                | Image Clustering using Moving K-mean Clustering | 71 |
| 3.4.1.3                                | Image Filtering using Median Filter             | 74 |
| 3.4.1.4                                | ROI Color Retrieval                             | 75 |

|         |  |     |
|---------|--|-----|
| 3.4.1.5 | Manual Segmentation using Adobe Photoshop                                | 77  |
| 3.4.2   | Segmentation of Red Blood Cells  | 78  |
| 3.4.2.1 | Seeded Region Growing Area Extraction Algorithm                          | 80  |
| 3.4.2.2 | Removal of Red Blood Cell that Touched the Image Border                  | 82  |
| 3.4.3   | Region Filling in Red Blood Cells  | 83  |
| 3.4.4   | Edge Detection Using Laplacian Operator                                  | 85  |
| 3.4.5   | Evaluation of the RBC Segmentation                                       | 87  |
| 3.5     | Morphological Feature Extraction for Red Blood Cells                     | 87  |
| 3.5.1   | Simple Shape based Features  | 88  |
| 3.5.2   | Color based Features   | 89  |
| 3.5.3   | Complex Shape based Features   | 91  |
| 3.6     | Feature Selection and Reduction Methods for Red Blood Cells              | 94  |
| 3.7     | Classification of Blood Cells Using Multilayer Perceptron Neural Network | 96  |
| 3.8     | Evaluation of Classification Performances                                | 99  |
| 3.9     | Summary  | 101 |

## **CHAPTER 4: RESULTS & DISCUSSION**

|       |  |     |
|-------|--|-----|
| 4.1   | Introduction                           | 103 |
| 4.2   | Contrast Enhancement Techniques        | 104 |
| 4.2.1 | Dark Contrast Enhancement Technique    | 112 |
| 4.2.2 | Bright Contrast Enhancement Technique  | 115 |
| 4.2.3 | Partial Contrast Enhancement Technique | 118 |

|         |   |     |
|---------|---|-----|
| 4.2.4   | Global Contrast Enhancement Technique                         | 121 |
| 4.2.5   | Local Contrast Enhancement Technique                          | 127 |
| 4.3     | Blood Cells Segmentation                                      | 130 |
| 4.3.1   | HSI Color Conversion  | 132 |
| 4.3.2   | Moving K-mean Clustering                                      | 135 |
| 4.3.4   | Segmented Blood Cells Images                                  | 140 |
| 4.3.5   | Segmentation Accuracy for Different Processing Techniques     | 147 |
| 4.4     | Red Blood Cells Segmentation                                  | 149 |
| 4.5     | Validation of the Proposed Image Segmentation Technique       | 153 |
| 4.6     | Data Features of Red Blood Cells                              | 157 |
| 4.7     | Feature Selection Methods                                     | 158 |
| 4.7.1   | Chi-Squared Filter  | 159 |
| 4.7.2   | InfoGain Filter   | 161 |
| 4.7.3   | CFS Filter  | 163 |
| 4.7.4   | PCA   | 164 |
| 4.8     | Classification of Red Blood Cells                             | 165 |
| 4.8.1   | Classification of Normal and Abnormal Red Blood Cells         | 166 |
| 4.8.1.1 | Classification Performance for All Features                   | 166 |
| 4.8.1.2 | Classification Performance for Selected Features              | 172 |
| 4.8.2   | Classification of Normal Blood and Types of Thalassemia Cells | 179 |
| 4.8.2.1 | Classification Performance for All Features                   | 179 |
| 4.8.2.2 | Classification Performance for Selected Features              | 181 |

|   |  |            |
|---|--|------------|
| 4.8.3   | Classification Performances of the Proposed Intelligent Screening<br>Technique for Thalassemia Based on the Blood Cells Extracted Features | 183        |
| 4.9   | Summary  | 187        |
| <b>CHAPTER 5: CONCLUSION &amp; FUTURE RECOMMENDATIONS</b> |  |            |
| 5.1   | Conclusion   | 189        |
| 5.2   | Research Contributions   | 191        |
| 5.3   | Recommendations for Future Work  | 192        |
|   | <b>REFERENCES</b>  | <b>194</b> |
|   | <b>APPENDIX</b>  | <b>207</b> |
|   | <b>LIST OF PUBLICATIONS</b>  | <b>214</b> |



## LIST OF TABLES

| NO   | PAGE |
|--|------|
| 2.1 Clinical Characteristic of $\alpha$ -thalassemia   | 13   |
| 2.2 Clinical Characteristic of $\beta$ -thalassemia  | 13   |
| 3.1 The first classification process of the blood cell   | 98   |
| 3.2 The second classification process of the blood cell  | 98   |
| 3.3 Architecture and training parameters of MLP network  | 99   |
| 3.4 Interpretation of a confusion matrix for classification evaluation   | 101  |
| 4.1 Assigation of pixels value for each number of clusters   | 136  |
| 4.2 The results of different processing techniques for an image  | 146  |
| 4.3 Average segmentation accuracies for 40 blood images  | 148  |
| 4.4 Segmentation accuracy of the segmented RBC images.   | 156  |
| 4.5 Overall Feature Set of RBC   | 158  |
| 4.6 The results of the feature selection processes   | 159  |
| 4.7 The selected features from the Chi-Squared filter  | 161  |
| 4.8 The selected features from the InfoGain filter   | 162  |
| 4.9 The selected features for RBC classification using CFS filter.   | 163  |
| 4.10 Proportion of input data for RBC classification.  | 165  |
| 4.11 The classification performances of three training algorithms for the classification between normal and abnormal RBC.                      | 171  |
| 4.12 The classification performances of four feature selection methods for the classification between normal and abnormal RBC.                 | 178  |
| 4.13: The classification performances of three training algorithms for the classification between normal RBC and the types of thalassemia.     | 180  |
| 4.14 The classification performances of four feature selection methods for the classification between normal RBC and the types of thalassemia. | 181  |

|      |   |     |
|------|---|-----|
| 4.15 | Classification performances for thalassemia screening based on individual RBC for MLP_LM network. | 183 |
| 4.16 | The statistical measures in RBC classification for individual blood sample.                       | 185 |

© This item is protected by original copyright

## LIST OF FIGURES

| NO   | PAGE |
|--|------|
| 1.1 Global Distribution of Haemoglobinopathies (World Health Organization, 2011)                                 | 2    |
| 2.1 Inheritance patterns of thalassemia disease (Nasser, 2013)   | 9    |
| 2.2 Inheritance patterns of thalassemia disease (Nasser, 2013)   | 9    |
| 2.3 Structure of red blood cell (Silverthorn, 2015)  | 11   |
| 2.4 Prevalence rate of thalassemia patients per 100,000 populations by states in Malaysia, 2009 (Ibrahim, 2012). | 15   |
| 2.5 Normal Blood Smear Image   | 16   |
| 2.6 Hypochromic Microcytic Anaemia Blood Smear Image   | 17   |
| 2.7 Cropped samples of available blood smear images showing different qualitative characteristics.               | 17   |
| 2.8 Relative migrations of haemoglobin variants on alkaline electrophoresis (Fischbach & Dunning, 2009)          | 20   |
| 2.9 Breast cancer images (Sheshadri & Kandaswamy, 2005)  | 23   |
| 2.10 Digitization process (Gonzalez & Woods, 2007)   | 24   |
| 2.11 Grayscale Image with Pixel Information<br>Pixel coordinates: (1364, 1366), Number of pixel: [158]           | 25   |
| 2.12 RGB Image with Pixel Information<br>Pixel coordinates: (1364, 1366), Number of RGB pixels: [244 84 70]      | 25   |
| 2.13 RGB color model cube (Séquin, 2011)   | 26   |
| 2.14 HSI color space (Gasparri, Bouchet, Abras, et al., 2011)  | 27   |
| 2.15 Dark Contrast Process   | 29   |
| 2.16 Bright Contrast Process   | 30   |
| 2.17 Partial Contrast Process  | 31   |
| 2.18 A feed-forward network  | 48   |
| 2.19 A feedback network  | 48   |

|      |  |     |
|------|--|-----|
| 2.20 | A generic model of MLP classifier  | 49  |
| 3.1  | The development of the screening procedure for thalassemia disease   | 62  |
| 3.2  | A set of digital camera mounted on Nikon light microscope with controllable LCD monitor                            | 64  |
| 3.3  | Original thalassemia image   | 65  |
| 3.4  | The intensity histogram of original thalassemia image  | 66  |
| 3.5  | The proposed image segmentation procedures for blood images  | 67  |
| 3.6  | The ROI retrieval process  | 75  |
| 3.7  | Enhancement techniques and color conversion components involved in the blood cell segmentation                     | 76  |
| 3.8  | Manual segmentation process  | 78  |
| 3.9  | Final result of manual segmentation  | 78  |
| 3.10 | Input images for the process of SRGAE technique  | 82  |
| 3.11 | The effect of RBC removal on $\beta$ -thalassemia blood image  | 83  |
| 3.12 | The effect of filling the holes in $\beta$ -thalassemia blood image  | 84  |
| 3.13 | Laplacian operator convolution mask  | 86  |
| 3.14 | The first process of RBC classification  | 96  |
| 3.15 | The second process of RBC classification   | 97  |
| 4.1  | Original images of $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.       | 106 |
| 4.2  | Intensity histograms for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood. | 107 |
| 4.3  | RGB histograms for $\alpha$ -thalassemia.  | 108 |
| 4.4  | RGB histograms for $\beta$ -thalassemia trait.   | 109 |
| 4.5  | RGB histograms for $\beta$ -thalassemia.   | 110 |
| 4.6  | RGB histograms for normal blood.   | 111 |

|      |   |     |
|------|---|-----|
| 4.7  | Dark contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.                           | 113 |
| 4.8  | Intensity histograms of dark contrast enhancement result for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.    | 114 |
| 4.9  | Bright contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.                         | 117 |
| 4.10 | Intensity histograms of bright contrast enhancement result for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.  | 118 |
| 4.11 | Partial contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.                        | 120 |
| 4.12 | Intensity histogram of partial contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood. | 121 |
| 4.13 | Global contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.                         | 122 |
| 4.14 | Intensity histogram of global contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.  | 123 |
| 4.15 | Result for global contrast enhancement technique with various selected percentage values.   | 124 |
| 4.16 | Intensity histogram of the image after applying global contrast enhancement technique with various selected percentage values.                            | 125 |
| 4.17 | Result for global contrast enhancement technique with various selected percentage values.   | 125 |
| 4.18 | Intensity histogram of the image after applying global contrast enhancement technique with various selected percentage values.                            | 126 |
| 4.19 | Local contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.                          | 129 |
| 4.20 | Intensity histogram of local contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood.   | 129 |
| 4.21 | Dark contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood                            | 131 |
| 4.22 | Partial contrast enhancement results for $\alpha$ -thalassemia, $\beta$ -thalassemia trait, $\beta$ -thalassemia and normal blood                         | 132 |

|      |  |     |
|------|--|-----|
| 4.23 | HSI color conversion results from dark contrast images   | 133 |
| 4.24 | HSI color conversion results from partial contrast images  | 134 |
| 4.25 | Resultant images from clustering & filtering Hue image of dark contrast in Figure 4.23(j) with various numbers of clusters           | 137 |
| 4.26 | Resultant images from clustering & filtering Hue image of partial contrast in Figure 4.24(j) with various numbers of clusters        | 137 |
| 4.27 | Resultant images from clustering & filtering Saturation image of dark contrast in Figure 4.23(k) with various numbers of clusters    | 138 |
| 4.28 | Resultant images from clustering & filtering Saturation image of partial contrast in Figure 4.24(k) with various numbers of clusters | 138 |
| 4.29 | Resultant images from clustering & filtering Intensity image of dark contrast in Figure 4.23(l) with various numbers of clusters     | 139 |
| 4.30 | Resultant images from clustering & filtering Intensity image of partial contrast in Figure 4.24(l) with various numbers of clusters  | 139 |
| 4.31 | Retrieval pixels of the images from Figure 4.25 with various numbers of clusters   | 141 |
| 4.32 | Retrieval pixels of the images from Figure 4.26 with various numbers of clusters   | 142 |
| 4.33 | Retrieval pixels of the images from Figure 4.27 with various numbers of clusters   | 142 |
| 4.34 | Retrieval pixels of the images from Figure 4.28 with various numbers of clusters   | 143 |
| 4.35 | Retrieval pixels of the images from Figure 4.29 with various numbers of clusters   | 144 |
| 4.36 | Retrieval pixels of the images from Figure 4.30 with various numbers of clusters   | 144 |
| 4.37 | The range of pixels  | 145 |
| 4.38 | Resultant images from 3 clusters blood cells segmentation based on Intensity component of dark contrast enhancement technique.       | 150 |
| 4.39 | The resultant images for each procedure in RBC segmentation  | 151 |
| 4.40 | Results of the segmented RBC images  | 152 |

|      |   |     |
|------|---|-----|
| 4.41 | (a),(d) Manually segmented RBC, (b),(e) Automatically segmented RBC and (c),(f) Ghost image of the segmented RBC for $\alpha$ -thalassemia and $\beta$ -thalassemia trait images, respectively. | 154 |
| 4.42 | (a),(d) Manually segmented RBC, (b),(e) Automatically segmented RBC and (c),(f) Ghost image of the segmented RBC for $\beta$ -thalassemia and normal blood images, respectively.                | 155 |
| 4.43 | The features weight versus the number of features for the classification between normal and abnormal RBC using Chi-Square filter.   | 160 |
| 4.44 | The features weight versus the number of features for the classification between normal and the types of thalassemia using Chi-Square filter.   | 160 |
| 4.45 | The features weight versus the number of features for the classification between normal and abnormal RBC using InfoGain filter.   | 161 |
| 4.46 | The features weight versus the number of features for the classification between normal and the types of thalassemia using InfoGain filter.   | 162 |
| 4.47 | The graph of the total residual power versus principal components   | 164 |
| 4.48 | Validation accuracy versus training epoch for the classification between normal and abnormal RBC using MLP_LM network.  | 167 |
| 4.49 | Validation accuracy versus hidden node for classification between normal and abnormal RBC using MLP_LM network.   | 167 |
| 4.50 | Validation accuracy versus training epoch for the classification between normal and abnormal RBC using MLP_BR network.  | 168 |
| 4.51 | Validation accuracy versus hidden node for classification between normal and abnormal RBC using MLP_BR network.   | 169 |
| 4.52 | Validation accuracy versus training epoch for the classification between normal and abnormal RBC using MLP_SCG network.   | 170 |
| 4.53 | Validation accuracy versus hidden node for classification between normal and abnormal RBC using MLP_SCG network.  | 170 |
| 4.54 | Validation accuracy versus training epoch of the classification between normal and abnormal RBC for MLP_LM network by using Chi-Squared results.  | 173 |
| 4.55 | Validation accuracy versus hidden nodes of the classification between normal and abnormal RBC for MLP_LM network by using Chi-Squared results.  | 173 |

|      |   |     |
|------|---|-----|
| 4.56 | Validation accuracy versus training epoch of the classification between normal and abnormal RBC for MLP_LM network by using InfoGain results. | 174 |
| 4.57 | Validation accuracy versus hidden nodes of the classification between normal and abnormal RBC for MLP_LM network by using InfoGain results.   | 175 |
| 4.58 | Validation accuracy versus training epoch of the classification between normal and abnormal RBC for MLP_LM network by using CFS results.      | 176 |
| 4.59 | Validation accuracy versus hidden nodes of the classification between normal and abnormal RBC for MLP_LM network by using CFS results.        | 176 |
| 4.60 | Validation accuracy versus training epoch of the classification between normal and abnormal RBC for MLP_LM network by using PCA results.      | 177 |
| 4.61 | Validation accuracy versus hidden nodes of the classification between normal and abnormal RBC for MLP_LM network by using PCA results.        | 177 |

© This item is protected by original copyright



## LIST OF ABBREVIATIONS

|          |  |
|----------|--|
| ALL      | Acute Lymphocytic Leukemia                 |
| AML      | Acute Myelogenous Leukemia                 |
| ANN      | Artificial Neural Network                  |
| BMP      | Bitmap                                     |
| BR       | Bayesian Regulation                        |
| CBC      | Complete Blood Count                       |
| CFS      | Correlation Feature Selection              |
| DNA      | Deoxyribonucleic Acid                      |
| FN       | False Negative                             |
| FP       | False Positive                             |
| Hb       | Haemoglobin                                |
| Hct      | Hematocrit                                 |
| HSI      | Hue, Saturation, Intensity                 |
| HUSM     | Hospital Universiti Sains Malaysia         |
| InfoGain | Information Gain                           |
| LM       | Levenberg-Marquardt                        |
| LVQ      | Learning Vector Machine                    |
| MCV      | Mean Corpuscular Volume                    |
| MCH      | Mean Corpuscular Haemoglobin               |
| MCHC     | Mean Corpuscular Haemoglobin Concentration |
| MLP      | Multilayer Perceptron                      |
| PCA      | Principle Component Analysis               |
| PSNR     | Peak-signal-to-noise-ratio                 |

|       |                                     |
|-------|-------------------------------------|
| RBC   | Red Blood Cell                      |
| RDW   | Red cell Distribution Width         |
| RGB   | Red, Green, Blue                    |
| ROI   | Region of Interest                  |
| SCG   | Scale Conjugate Gradient            |
| SFS   | Sequential Forward Selection        |
| SRGAE | Seed Region Growing Area Extraction |
| SVM   | Support Vector Machine              |
| TN    | Truth Negative                      |
| TP    | Truth Positive                      |
| WBC   | White blood cell                    |
| WHO   | World Health Organization           |

© This item is protected by original copyright

## LIST OF SYMBOLS

|                             |   |
|-----------------------------|---|
| $A_c$                       | Area of segmented RBC                               |
| $SF_b$                      | Bright stretching factor                            |
| $B_c$                       | Blue pixels of RBC                                  |
| $\mu_{pq}$                  | Central Moment                                      |
| $f(x,y)$                    | Color level of image in the location of [x,y]       |
| $in(x,y)$                   | Color level for input pixel                         |
| $q_k$                       | Color level of input pixel                          |
| $out(x,y)$                  | Color level for output pixel                        |
| $P_k$                       | Color level of output pixel                         |
| $SF_d$                      | Dark stretching factor                              |
| dL                          | Decilitre   |
| $max$                       | Desired maximum color levels in output image        |
| $min$                       | Desired minimum color levels in output image        |
| fL                          | Femtolitre (a fraction of one-millionth of a litre) |
| $\nabla V(\underline{x})$   | Gradient  |
| gm                          | Grams   |
| $G_c$                       | Green pixels of RBC                                 |
| $\nabla^2 V(\underline{x})$ | Hessian matrix                                      |
| HN                          | Hidden node   |
| $[i,j]$                     | Image pixel location                                |
| $f[i,j]$                    | Image pixel value                                   |
| $I$                         | Input node  |

|                  |   |
|------------------|---|
| $\nabla^2 f$     | Laplacian operator                            |
| $minTH$          | Lower threshold value                         |
| $\mu$            | Marquardt adjustment parameter                |
| $maxB$           | Maximum color level of blue                   |
| $maxG$           | Maximum color level of green                  |
| $maxR$           | Maximum color level of red                    |
| $maxRGB$         | Maximum color level of RGB                    |
| $f_{max}$        | Maximum color level values in the input image |
| $\overline{B_c}$ | Mean blue of RBC                              |
| $\overline{G_c}$ | Mean green of RBC                             |
| $\overline{R_c}$ | Mean red of RBC                               |
| $\overline{I_c}$ | Mean intensity of RBC                         |
| $\mu m$          | Micrometer                                    |
| $minB$           | Minimum color level of blue                   |
| $minG$           | Minimum color level of green                  |
| $minR$           | Minimum color level of red                    |
| $minRGB$         | Minimum color level of RGB                    |
| $f_{min}$        | Minimum color level values in the input image |
| $m_{pq}$         | Moment  |
| $NminTH$         | New lower stretching value                    |
| $NmaxTH$         | New upper stretching value                    |
| $out_{rgb}(x,y)$ | New RGB pixel value                           |
| $\alpha$         | Normal alpha gene                             |
| $\beta$          | Normal beta gene                              |
| $\beta^0$        | No beta gene formation                        |

|                 |                                |
|-----------------|--------------------------------|
| $in_{RGB}(x,y)$ | Original RGB pixel value       |
| $R_c$           | Red pixels of RBC              |
| $\beta^+$       | Some beta gene formation occur |
| $\theta$        | Theta                          |
| $maxTH$         | Upper threshold value          |
| $\sigma^2 B_c$  | Variance blue of RBC           |
| $\sigma^2 G_c$  | Variance green of RBC          |
| $\sigma^2 R_c$  | Variance red of RBC            |
| $\sigma^2 I_c$  | Variance intensity of RBC      |

© This item is protected by original copyright

## Teknik-teknik Peningkatan Imej dan Pengekstrakan Ciri untuk Prosedur Saringan Talasemia Pintar

### ABSTRAK

Talasemia adalah penyakit darah yang diwarisi dan memberi kesan kepada pengeluaran sel darah merah (RBC) di dalam badan. Kekurangan RBC yang terdiri daripada 99% sel darah secara keseluruhan akan menjejaskan fungsi utama sebagai agen pembawa oksigen. Ciri morfologi RBC memainkan peranan penting dalam diagnosis penyakit tersebut. Buat masa ini, siasatan mikroskopik dijalankan secara manual dalam mengenal pasti kehadiran sel talasemia oleh pakar hematologi melalui mikroskop cahaya. Walau bagaimanapun, prosedur manual ini boleh menghasilkan keputusan yang kurang tepat, memerlukan tenaga kerja yang intensif dan memakan masa kerana ia terlalu bergantung kepada pengalaman dan kemahiran pakar hematologi. Oleh itu, objektif utama penyelidikan ini adalah membangunkan prosedur saringan talasemia pintar berdasarkan sampel darah. Satu kaedah yang efisien telah dibina dengan menggunakan teknik pemrosesan imej termasuklah peningkatan imej, peruasan imej dan pengekstrakan ciri bagi mendapatkan peruasan RBC sepenuhnya. Terdapat lima jenis teknik peningkatan kontras yang telah diaplikasikan ke atas imej asal RGB secara berasingan dan dua teknik dipilih untuk digunakan dalam kaedah cadangan iaitu teknik peningkatan kontras gelap dan separa. Seterusnya, peruasan sel darah diteruskan dengan penukaran kepada model warna HSI sebelum imej itu diproses dengan menggunakan kaedah kelompok purata- $k$  bergerak dan kaedah penapisan median. Prestasi peruasan bagi sel darah yang baik telah diperolehi daripada kombinasi kaedah peningkatan kontras gelap dan komponen keamatan. Kemudian, imej RBC yang telah diruas sepenuhnya diperolehi selepas komponen lain yang tidak dikehendaki seperti RBC yang bersentuhan, WBC dan platelet berjaya dibuang dengan menggunakan kaedah pertumbuhan rantauan. Tiga ciri utama diekstrak daripada RBC individu berdasarkan bentuk ringkas, warna dan bentuk kompleks. Pengekstrakan ciri ini telah menghasilkan 31 ciri-ciri dan dijadikan sebagai data masukan dalam rangkaian perseptron berbilang lapisan. Terdapat dua proses klasifikasi RBC dalam kajian ini. Proses klasifikasi pertama adalah mengklasifikasikan jenis RBC kepada normal atau tidak normal (talasemia). Proses klasifikasi yang kedua pula adalah mengklasifikasikan sel normal dan tiga jenis talasemia yang dinamakan  $\alpha$ -talasemia,  $\beta$ -talasemia trait dan  $\beta$ -talasemia. Algoritma *Levenberg-Marquardt* telah menghasilkan prestasi klasifikasi yang terbaik setelah dibandingkan dengan algoritma regulasi *Bayesian* dan algoritma skala konjugat kecerunan. Kemudian, ciri-ciri yang telah diekstrak itu akan diproses dengan menggunakan empat kaedah pemilihan ciri secara berasingan iaitu khi kuasa dua, kenaikan maklumat, korelasi ciri pemilihan dan analisis komponen utama. Peratus ketepatan klasifikasi yang tertinggi diperolehi bagi kedua-dua klasifikasi RBC dengan menggunakan ciri-ciri yang telah terpilih daripada penapis khi kuasa dua. Hal ini disebabkan oleh ciri-ciri terpenting telah digunakan sebagai data masukan. Prestasi klasifikasi yang optimum telah dicapai dengan kejituan pengesahsahihan dan kejituan ujian masing-masing dengan 98% dan 96.8% untuk klasifikasi yang pertama. Manakala, kejituan pengesahsahihan dan kejituan ujian masing-masing dengan 95.7% dan 94.4% diperolehi dalam klasifikasi yang kedua. Maka, pembangunan teknik peningkatan imej dan teknik pengekstrakan ciri RBC dalam penyelidikan ini memberikan satu alternatif yang cekap dalam menganalisis dan mengklasifikasikan sampel darah.

## Image Enhancement and Feature Extraction Techniques for Intelligent Thalassemia Screening Procedures

### ABSTRACT

Thalassemia is an inherited blood disease that effects the production of red blood cells (RBC) in the body. The deficiency of RBC that constitutes 99% of blood cells will affect their main function as oxygen carrier. The morphological features of RBC play a crucial role in medical diagnosis. Currently, the microscopic investigation to identify the present of any thalassemia cells is performed manually by haematologists through visual identification under a light microscope. However, the manual procedure yields inaccurate results, labour-intensive and time-consuming since it is highly dependent on the haematologists experience and skill. Thus, the main objective of this research is to develop an intelligent thalassemia screening procedure based on blood samples. Essentially an efficient method using the image processing techniques including image enhancement, segmentation and feature extraction have been constructed in order to obtain a fully segmented RBC. There are five contrast enhancement techniques have been applied to the original RGB image, separately and two techniques were selected to be used in the proposed procedure that are Dark Contrast and Partial Contrast techniques. Then, the segmentation of blood cells proceed with the conversion of HSI color space before the image being processed using moving  $k$ -mean clustering and median filter techniques. Good segmentation performance for blood cell has been obtained from the combination of dark contrast technique and intensity component. Then, fully segmented RBC image was obtained after the unwanted components such as overlapping RBC, WBC and platelet were successfully removed using seed region growing technique. Next, three main features were extracted from the individual RBC that are simple shape, color and complex shape based features. The features extraction has produced 31 features that have been fed as inputs to the Multilayer Perceptron (MLP) network. There are two processes of RBC classification have been performed in this research. The first process was carried out to classify the type of RBC into normal or abnormal (thalassemia). Then, the second process was continued to classify the normal blood and three types of thalassemia cells namely  $\alpha$ -thalassemia,  $\beta$ -thalassemia trait and  $\beta$ -thalassemia. The Levenberg-Marquardt (LM) algorithm produces the best classification performance compared to Bayesian Regulation (BR) and Scale Conjugate Gradient (SCG) algorithms. Next, the extracted features were processed using four feature selection methods that are Chi-Squared, Information Gain, Correlation-based Feature Selection and Principal Component Analysis, separately. The highest classification accuracy was obtained for both RBC classifications using the selected features from Chi-Squared filter. This is due to selected significant features were used as inputs in the network. The optimal classification performance has been achieved with validation accuracy of 98% and testing accuracy of 96.8% for first classifier (11 features). While, validation and testing accuracies of 95.7% and 94.4%, respectively have been achieved in the second classifier (14 features). Therefore, the development of image enhancement and feature extraction techniques in this research provides an efficient alternative in analyzing and classifying the blood sample.

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Haemoglobinopathies, or haemoglobin disorder, is one of the inherited blood diseases characterized by the presence of abnormal haemoglobin in the blood. The cause of this abnormality are due to the inheritance of mutated genes that interfere with the formation of haemoglobin molecules in the red blood cells with its primary function as a transporter of oxygen to the whole body (Ferry, 1923). If the oxygen has not been transported sufficiently, it will cause further damage or functional disability of certain organs and tissues.

World Health Organization (WHO) estimates that approximately 2.9% of thalassemia and 2.3% of sickle cell disease are the carriers of the abnormal haemoglobin condition (Memish, Owaidah & Saeedi, 2011). In fact, the prevalence of these diseases is greater than cystic fibrosis and haemophilia (Anionwu & Atkin, 2001). Over 300,000 to 500,000 babies are died each year due to the severe haemoglobin disorders (WHO, 2011). As a result of global migration patterns, the diseases are spreading aggressively throughout the countries including Malaysia. Figure 1.1 shows the global distribution of haemoglobinopathies in terms of births per 1,000 affected infants.