DESIGN OF A SCREENING SYSTEM FOR ACUTE LEUKEMIA CELLS BASED ON BONE MARROW SAMPLES

ELSIE USUN FRANCIS

UNIVERSITI MALAYSIA PERLIS

2012

DESIGN OF A SCREENING SYSTEM FOR ACUTE LEUKEMIA CELLS BASED ON BONE MARROW SAMPLES

by

ELSIE USUN FRANCIS
(0931310353)

A thesis submitted in fulfillment of the requirements for the degree of Master of Science (Biomedical Electronic Engineering)

School of Mechatronic Engineering
UNIVERSITI MALAYSIA PERLIS

2012

UNIVERSITI MALAYSIA PERLIS

DECLARATION OF THESIS

Author's full name : ELSIE USUN FRANCIS

Date of birth : 17 SEPTEMBER 1986

Title : DESIGN OF A SCREENING SYSTEM FOR ACUTE LEUKEMIA CELLS BASED ON BONE MARROW SAMPLES

Academic Session : 2009-2012

I hereby declare that the thesis becomes the property of Universiti Malaysia Perlis (UniMAP) and to be placed at the library of UniMAP. This thesis is classified as :

CONFIDENTIAL ☐ (Contains confidential information under the Official Secret Act 1972)*

RESTRICTED ☐ (Contains restricted information as specified by the organization where research was done)*

OPEN ACCESS ☐ I agree that my thesis is to be made immediately available as hard copy or on-line open access (full text)

I, the author, give permission to the UniMAP to reproduce this thesis in whole or in part for the purpose of research or academic exchange only (except during a period of _____ years, if so requested above).

Certified by:

_____ _____

SIGNATURE **SIGNATURE OF SUPERVISOR**

(NEW IC NO. / PASSPORT NO.) **NAME OF SUPERVISOR**

Date :_____ Date : _____

NOTES : * If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentially or restriction.

ACKNOWLEDGEMENT

The completion of this research and thesis writing is done with great support and help from a number of people. Firstly, I wish to express my greatest gratitude especially to my honorable supervisor, Prof. Dr. Mohd. Yusoff Mashor for his kind offers of advice and guidance in any ways possible to make this research a great achievement for me. I would also like to thank my co-supervisors, Assoc. Prof. Dr. Rosline Hassan and Miss Azian Azamimi Abdullah for their constructive guidance and support all through the way.

To my beloved parents and family who are always there to support me mentally and emotionally, I am deeply indebted. Their endless love, motivation and encouragement have given me strength to bring about this research and thesis writing. Thank you so much.

Finally, special thanks to all my fellow friends in the Autonomous Research Cluster, UniMAP especially to Madam Nor Hazlyna Harun, Madam Rafikha Aliana A. Raof, Madam Robiyanti Adollah, Madam Noor Sara Mashitoh Binti Mohamad Hanif, Miss Lim Huey Nee, Miss Nurul Hazwani Abdul Halim, Miss Norhayati Binti Mohd Nazid and Madam Intan Maisarah Abdul Rahim. Not forgetting, many thanks to the obliging staffs in Hematologist Department, HUSM especially to Miss Selamah Ghazali and all who has involved deliberately or inadvertently in the completion of this research. This research will not be successfully through without the help of all.

Thank you very much.

TABLE OF CONTENT

	PAGE
THESIS DECLARATION	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENT	iii
LIST OF TABLES	viii
LIST OF FIGURES	xi
LIST OF ABBREVIATION	xiv
LIST OF SYMBOL	xvii
ABSTRAK	xx
ABSTRACT	xxii

CHAPTER 1: INTRODUCTION

1.1	Introduction	1
1.2	Problem Statement	2
1.3	State-of-the-Art	4
1.4	Objectives of Research	6
1.5	Scope of Research	6
1.6	Overview of Thesis	7

CHAPTER 2: LITERATURE REVIEW

2.1	Introduction	10
2.2	Leukemia	11

2.2.1 Bone Marrow 12

2.2.2 Leukemia Risk Factors and Signs and Symptoms 16

2.2.3 Diagnosing Leukemia 17

2.2.3.1 Physical Exam 18

2.2.3.2 Blood Tests 18

2.2.3.3 Bone Marrow Tests 19

2.2.3.4 Other Tests 21

2.2.3.5 Drawbacks of Conventional Tests 23

2.3 Digital Image 25

2.3.1 RGB Image 25

2.3.2 HSI Color Space 26

2.3.3 Digital Image Processing on Medical Images 26

2.3.3.1 Image Enhancement 27

2.3.3.2 Image Segmentation 34

2.3.3.3 Features Extraction 42

2.4 Classification 45

2.4.1 Artificial Neural Network 47

2.4.1.1 Background of Multilayered Perceptron 47

2.4.1.2 Applications of Multilayered Perceptron on Medical Images 52

2.4.2 Instance-based Learning Algorithm 53

2.4.2.1 Background of K-nearest Neighbor 53

2.4.2.2 Applications of K-nearest Neighbor on Medical Images 55

2.5 Existing Screening System for Leukemia 56

2.6 Summary 59

CHAPTER 3: METHODOLOGY

3.1 Introduction 60

3.2 Image Acquisition 62

3.3 Digital Image Processing on Bone Marrow Slide Images 63

3.3.1 Image Enhancement Techniques 65

3.3.2 Image Segmentation of White Blood Cells 68

3.3.2.1 Computation of Saturation Component 70

3.3.2.2 Thresholding using Saturation Component 71

3.3.2.3 Median Filtering 72

3.3.2.4 Removing Unwanted Regions using Seed-based Region Growing (SBRG) Algorithm 73

3.3.2.5 Separation of Overlapped White Blood Cells using Watershed Algorithm 74

3.4 Features Extraction of White Blood Cells 75

3.4.1 Morphological Features of White Blood Cells 76

3.4.2 Features Extraction Procedure for White Blood Cells 79

3.5 Screening and Classification 80

3.5.1 Structure Selected for MLP classifier 81

3.5.1.1 MLP Structure 82

3.5.1.2 MLP Input Data Division 84

3.5.1.3 Selection of Suitable Number of Hidden Neurons 85

3.5.1.4 Training Parameters of MLP networks 86

3.5.1.5 Methodology for Classification of Bone Marrow 87

Slide Images using Multilayer Perceptron

3.5.2 *k*-Nearest Neighbor Classifier 88

3.5.2.1 Parameters of *k*-NN 89

3.5.2.2 *k*-NN Input Data Division 90

3.5.2.3 Methodology for Classification of Bone Marrow Slide Images using *k*-NN 92

3.6 Summary 92

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction 94

4.2 Implementation of Image Enhancement Techniques 95

4.2.1 Original Images 96

4.2.2 Partial Contrast Stretching 99

4.2.3 Dark Contrast Stretching 105

4.3 Image Segmentation of White Blood Cells 111

4.3.1 Computation of Saturation Component 111

4.3.2 Thresholding 118

4.3.3 Median Filter 130

4.3.4 SBRG 134

4.4 Separation of Overlapped White Blood Cells using Watershed Algorithm 136

4.5 Screening and Classification of Bone Marrow Slide Images 146

4.5.1 Multilayer Perceptron Network 147

4.5.1.1 MLP network (1A) 147

4.5.1.2 MLP network (1B) and (1C) 152

4.5.2 *k*-Nearest Neighbor 161

4.5.2.1 *k*-NN classifier (2A) 161

4.5.2.2 *k*-NN classifier (2B) and (2C) 162

4.5.3 Analysis of Classification performance between MLP and *k*-NN 164

4.5.3.1 Standard Classifiers 164

4.5.3.2 Hierarchical Classifiers 165

4.6 Conclusion 165

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion 168

5.2 Future Work 171

REFERENCES 172

APPENDIX

Appendix A Results for Enhancement Techniques on Bone Marrow Slide Images 179

Appendix B Results for Image Segmentation on Bone Marrow Slide Images 187

Appendix C Results for Segmentation Accuracy in terms of WBC Count with SBRG 191

Appendix D Analysis of MLP Networks Performance with Seed Number of 0 to 4 200

LIST OF PUBLICATIONS 213

LIST OF AWARDS 214

LIST OF TABLES

		PAGE
Table 3.1	Morphological Features of white blood cell for normal and abnormal bone marrow	78
Table 3.2	Input Data for MLP	84
Table 3.3	Output nodes for MLP network (1A)	87
Table 3.4	Output nodes for MLP network (1B)	87
Table 3.5	Output nodes for MLP network (1C)	87
Table 3.6	Input Data for KNN	91
Table 4.1	Segmentation accuracy (WBC count) for 700 images	135
Table 4.2	Best accuracy rate for each seed number (MLP network (1A))	148
Table 4.3	Performance Analysis of Individual Feature (Acc %) (1A)	150
Table 4.4	Accuracy Comparison of combined features–(1A)	151
Table 4.5	Best accuracy rate for each seed number (MLP network (1B))	153
Table 4.6	Sensitivity (SP) and Specificity (SE) of MLP network (1B)	151
Table 4.7	Best accuracy rate for each seed number (MLP network (1C))	154
Table 4.8	Performance analysis of individual feature (Acc %) (1B)	156
Table 4.9	Performance analysis of individual feature (Acc %) (1C)	157
Table 4.10	Accuracy Comparison of combined features – LM (1B)	159
Table 4.11	Accuracy Comparison of combined features – LM (1C)	159
Table 4.12	Average accuracy of hierarchical MLP (1B) and (1C)	157
Table 4.13	Classification performance of *k*-NN Classifier (2A) with different value of k	162
Table 4.14	Classification performance of *k*-NN Classifier (2B) and (2C) with	163

different value of k

Table 4.15	Average accuracy of hierarchical *k*-NN (2B) and (2C)	163
Table C.1	Segmentation Accuracy for Normal Bone Marrow Slide Images	191
Table C.2	Segmentation Accuracy for Abnormal M1 Bone Marrow Slide Images	192
Table C.3	Segmentation Accuracy for Abnormal M2 Bone Marrow Slide Images	193
Table C.4	Segmentation Accuracy for Abnormal M3 Bone Marrow Slide Images	194
Table C.5	Segmentation Accuracy for Abnormal M4 Bone Marrow Slide Images	195
Table C.6	Segmentation Accuracy for Abnormal M5 Bone Marrow Slide Images	197
Table C.7	Segmentation Accuracy for Abnormal M7 Bone Marrow Slide Images	198
Table D.1	Performance of MLP network (1A) with seed number = 0	200
Table D.2	Performance of MLP network (1A) with seed number = 1	201
Table D.3	Performance of MLP network (1A) with seed number = 2	201
Table D.4	Performance of MLP network (1A) with seed number = 3	202
Table D.5	Performance of MLP network (1A) with seed number = 4	203
Table D.6	Performance of MLP network (1B) with seed number = 0	204
Table D.7	Performance of MLP network (1B) with seed number = 1	205
Table D.8	Performance of MLP network (1B) with seed number = 2	206
Table D.9	Performance of MLP network (1B) with seed number = 3	206
Table D.10	Performance of MLP network (1B) with seed number = 4	207
Table D.11	Performance of MLP network (1C) with seed number = 0	208

Table D.12 Performance of MLP network (1C) with seed number = 1 209

Table D.13 Performance of MLP network (1C) with seed number = 2 210

Table D.14 Performance of MLP network (1C) with seed number = 3 211

Table D.15 Performance of MLP network (1C) with seed number = 4 211

LIST OF FIGURES

		PAGE
Figure 2.1	Marrow – Principal site for blood formation	13
Figure 2.2	Stem Cell maturation	14
Figure 2.3	Cells of granulocytic series at various stages of maturation	15
Figure 2.4	Bone marrow images with Wright stain	21
Figure 2.5	Partial Contrast Stretching	32
Figure 2.6	Dark Contrast Stretching	33
Figure 2.7	Image histogram	37
Figure 2.8	3x3 neighborhood kernel median filter	39
Figure 2.9	Types of path for region growing process	40
Figure 2.10	Illustration of Watershed Algorithm	41
Figure 2.11	Schematic diagram of MLP model with 1 hidden layer	48
Figure 2.12	Voting mechanism for *k*-NN	54
Figure 3.1	A motorized stage microscope connected to a computer	63
Figure 3.2	The proposed digital image processing procedure employed on bone marrow slide images	65
Figure 3.3	Intensity histogram of bone marrow slide images	67
Figure 3.4	Segmentation process	69
Figure 3.5	Saturation component histogram	71
Figure 3.6	Structure of White Blood Cell	78
Figure 3.7	MLP structure	83
Figure 3.8	Classification with *k*-NN	89
Figure 4.1	Histogram of Original Normal_1	96

Figure 4.2	Histogram of Original Abnormal_1	97
Figure 4.3	Histogram of PCS Normal_1 with different threshold values	99
Figure 4.4	Histogram of PCS Abnormal_1 with different threshold values	101
Figure 4.5	Histogram of DCS Normal_1 with different threshold values	106
Figure 4.6	Histogram of DCS Abnormal_1 with different threshold values	107
Figure 4.7	Saturation component of PCS Normal_1	112
Figure 4.8	Saturation component of PCS Abnormal_1	113
Figure 4.9	Saturation component of DCS Normal_1	115
Figure 4.10	Saturation component of DCS Abnormal_1	116
Figure 4.11	Thresholded images of PCS Normal_1b and DCS Normal_1b	119
Figure 4.12	Thresholded images of PCS Abnormal_1b and DCS Abnormal_1b	121
Figure 4.13	Thresholded images of DCS Normal_1b	124
Figure 4.14	Thresholded images of DCS Abnormal_1b	126
Figure 4.15	Median Filtered Image of Thresholded DCS Normal_1b	131
Figure 4.16	Median Filtered Image of Thresholded DCS Abnormal_1b	132
Figure 4.17	SBRG of median filtered DCS Normal_1b	134
Figure 4.18	SBRG of median filtered DCS Abnormal_1b	134
Figure 4.19	Watershed Transformation of Abnormal_1	137
Figure 4.20	Watershed Transformation of Abnormal_2	140
Figure 4.21	Watershed Transformation of Abnormal_3	141
Figure 4.22	Watershed Transformation of Abnormal_4	143
Figure 4.23	Watershed Transformation of Abnormal_5	144

Figure 4.24	Accuracy of MLP network (1A)	149
Figure 4.25	Accuracy Comparison for MLP (1B) and (1C) using LM and BR	156
Figure 4.26	Accuracy Comparison for standard MLP and *k*-NN	164
Figure 4.27	Accuracy Comparison for hierarchical MLP and KNN	165

LIST OF ABBREVIATIONS

AHE	Adaptive histogram equalization
ALL	Acute lymphoblastic leukemia
AML	Acute myeloblastic leukemia
ANN	Artificial neural network
ASBRG	Automatic seed based region growing
BG	Background
BP	Backpropagation
BR	Bayesian Regularization
CAD	Computer aided diagnosis
CBC	Complete blood count
CLAHE	Contrast-limiting adaptive histogram equalization
CLL	Chronic lymphoblastic leukemia
CML	Chronic myeloblastic leukemia
CT	Computed tomography
DBC	Differential blood count
DCS	Dark contrast Stretching
EKG	Electrocardiography
FAB	French American British
FCM	Fuzzy C-mean
FN	False negative
FP	False positive
GLCM	Gray level co-occurrence matrix
H^2MLP	Hierarchical hybrid multilayer perceptron

HE	Histogram equalization
HN	Hidden neuron
HSI	Hue Saturation Intensity
HSIL	High grade squamous intraepithelial lesion
HSV	Hue Saturation Value colour space
IBL	Instance-based learning
KNN	K-nearest neighbor
L1	Acute lymphoblastic leukemia subtype
L2	Acute lymphoblastic leukemia subtype
L3	Acute lymphoblastic leukemia subtype
LDA	Linear discriminant analysis
LM	Levenberg Marquardt
LSIL	Low grade squamous intraepithelial lesion
LVQ	Learning Vector Quantization
M0 – M7	Acute myeloblastic leukemia subtype
MAb	Monoclonal antibody
MLP	Multilayer Perceptron
PAS	Periodic acid-shiff
PCS	Partial contrast stretching
RBC	Red blood cell
RBF	Radial basis function
RGB	Red green blue color space
RVM	Relevance vector machine
SBRG	Seed based region growing
SE	Sensitivity

SP	Specificity
SRGFE	Seeded region growing features extraction
SVM	Support Vector Machine
TA	Training algorithm
TB	Tubercle Bacilli
TH	Threshold Value
TN	True negative
TP	True positive
WBC	White blood cell

LIST OF SYMBOLS

f_i'	Final values of a feature
f_i	Initial values of a feature
$\overline{R_n}$	Mean for red color value
$\overline{G_n}$	Mean for green color value
$\overline{B_n}$	Mean for blue color value
$\overline{VR_n}$	Variance for red color value
$\overline{VG_n}$	Variance for green color value
$\overline{VB_n}$	Variance for blue color value
$\overline{SR_n}$	Standard Deviation for red color value
$\overline{SG_n}$	Standard Deviation for green color value
$\overline{SB_n}$	Standard Deviation for blue color value
P_k	Color level of the output pixel
$T_j(p,q)$	Topological distance between p and q
$W_{shed}(f)$	Watershed lines
$d_{Chessboard}$	Distance Transform using 'chessboard'
$d_{Euclidean}$	Distance transform using 'euclidean'
[k]	Number of current region of interest
B	Value of blue
c	Circularity
$\underline{e}(\underline{x})$	Network error

f_{max}	Maximum color level values in the input image
f_{min}	Minimum color level values in the input image
G	Value of green
i_1, j_1	Coordinate of object (WBC) pixel
i_2, j_2	Coordinate of non-object (Background) pixel
$in(x, y)$	Color level of the input pixel.
$min\ \&\ max$	Desired color levels that determine color range of the output image
$maxTH$	Upper threshold value
$minTH$	Lower threshold value
n	Area of white blood cell
$NewmaxTH$	New upper stretching value
$NewminTH$	New lower stretching value
$NewTH$	Dark stretching factor
$out(x, y)$	Color level of the output pixel
p	Perimeter
q_k	Color level of the input pixel
R	Value of red
r	Radius
size [k]	Area of k
T	Threshold value for thresholding
TH	Threshold value for dark contrast stretching
$\mu(f_i)$	Mean of all the values of class that features belong to
$\sigma(f_i)$	Standard Deviation of all the values of class that features belong

	to
$CB(m_i)$	Catchment basin
$\nabla^2 V(\underline{x})$	Hessian matrix
I	Identity matrix
$J(\underline{x})$	Jacobian matrix
S	Saturation component
$f(x, y)$	Input image
$\nabla V(\underline{x})$	Gradient
$g(x, y)$	Output of thresholded image
$maxBlue$	Maximum blue level values
$maxGreen$	Maximum green level values
$maxRed$	Maximum red level values
$maxTH$	Average number of these maximum color level values for each color space
$minBlue$	Minimum blue level values
$minGreen$	Minimum green level values
$minRed$	Minimum red level values
$minTH$	Average number of these minimum color level values for each color space

MEREKABENTUK SISTEM UJIAN SARINGAN SEL LEUKEMIA AKUT

BERDASARKAN SAMPEL SUM - SUM TULANG

ABSTRAK

Kebolehan untuk mengesan kehadiran sel leukemia berdasarkan sampel sum - sum tulang dapat membantu para doktor untuk mengesahkan ujian saringan yang dibuat melalui darah. Walau bagaimanapun, imej slaid sum - sum tulang mempunyai beberapa kekangan seperti kehadiran kawasan latar belakang dan kontras imej yang rendah. Imej slaid sum-sum tulang adalah lebih baik jika kawasan latar belakang dan kesan kontras imej yang rendah dapat dikurangkan. Sehubungan itu, satu sistem pemprosesan imej bersama dengan keupayaan untuk mengelas imej sum – sum tulang telah dibangunkan bertujuan untuk mengurangkan kekangan – kekangan yang boleh mengganggu keputusan diagnosis leukemia yang dibuat secara manual. Terdapat dua jenis teknik peningkatan kontras yang telah diimplemen ke atas imej slaid sum - sum tulang. Teknik – teknik yang telah di implemen adalah algoritma tidak seluruh (partial contrast stretching) dan algoritma kontras gelap (dark contrast stretching). Walaupun, kedua - dua teknik peningkatan kontras yang dicadangkan telah memberikan keputusan yang baik ke atas imej slaid sum – sum tulang, algoritma kontras gelap telah dipilih untuk digunakan dalam penyelidikan ini. Proses penyingkiran latar belakang dan sel darah merah telah dimulakan dengan menggunakan komponen ketepuan daripada imej warna HSI. Kehadiran kawasan selain sel darah putih akan disingkir dengan menggunakan algoritma penapisan median. Selain daripada itu, penapisan median juga digunakan untuk mengisi ruang kosong (piksel putih) yang terdapat pada sel darah putih. Teknik peruasan kawasan pula digunakan untuk meruas, menyingkirkan kawasan – kawasan tertentu, serta mengekstrak ciri luas permukaan sel darah putih. Teknik ini mampu mengekalkan saiz dan bentuk asal sel darah putih yang terdapat di dalam imej slaid sum-sum tulang. Tidak seperti didalam darah, sel-sel yang terdapat di dalam kawasan sum – sum tulang saling bertindih di antara satu sama lain. Algoritma batas air (watershed algorithm) telah dilaksanakan ke atas sel-sel darah putih ini untuk memisahkan sel – sel tersebut. Selain daripada ciri luas permukaan, ciri - ciri geometrikal yang lain juga turut diekstrak daripada sel-sel darah putih seperti 'circularity', jejari dan perimeter sel. Ciri – ciri lain seperti warna juga diekstrak termasuk purata, sisihan piawai dan varians untuk setiap RGB komponen; merah, hijau dan biru. Ciri - ciri yang telah diekstrak ini akan dijadikan sebagai data masukan kepada rangkaian MLP dan *k*-NN untuk dikelaskan kepada imej normal, tidak normal jenis M3 dan tidak normal lain – lain. Algoritma pembelajaran Levenberg Marquardt (LM) dan Bayesian Regularization (BR) telah digunakan untuk menganalisis keupayaan rangkaian MLP untuk tujuan pengesanan. Rangkaian MLP (1A)-BR standard telah memperoleh kejituan yang paling tinggi berbanding rangkaian MLP (1A)-LM dan juga KNN (2A) untuk mengesan imej normal, tidak normal jenis M3 dan tidak normal lain – lain iaitu sebanyak 93.9% ke atas data ujian. Rangkaian MLP hierarki (1B) dan (1C) pula telah memperoleh kejituan sebanyak 99.8% ke atas data latihan dan 98.57% ke atas data ujian Secara keseluruhan, kesemua rangkaian MLP berupaya mengatasi *k*-NN untuk pengesanan imej sum – sum tulang dengan peratusan kejituan yang tinggi.

DESIGN OF A SCREENING SYSTEM FOR ACUTE LEUKEMIA CELLS BASED ON BONE MARROW SAMPLES

ABSTRACT

The capability to screen for leukemia based on bone marrow samples could facilitate the doctors in confirming the occurrence of leukemia from blood test. However, the images of the bone marrow slide have several drawbacks such as the appearance of unwanted regions and lack of contrast. The acquired images of the bone marrow slide could be better improved if these drawbacks are reduced. Due to these matters, a digital image processing system with classification capability is built up in this research which aims to reduce the drawbacks arise from manual screening of bone marrow slide. In this research, two enhancement techniques were used to improve the appearance of the acquired bone marrow slide images. These techniques include partial contrast stretching (PCS) and dark contrast stretching (DCS). Although both techniques produced good results, DCS has been chosen to be utilized in this research due to several reliable reasons. The elimination of unwanted regions leaving only the white blood cells (WBCs) in the bone marrow slide images is initiated by using the saturation component of the HSI (Hue, Saturation, Intensity) color space. Some noises (unwanted small particles) that still appeared were removed with median filter. Simultaneously, median filter fills 'holes' (white color pixel) which are enclosed within the WBCs. Subsequently, seed-based region growing (SBRG) algorithm is used to remove unwanted regions based on predefined criteria and at the same time extract the area of WBCs in the bone marrow slide images. SBRG technique is capable to maintain the original size and shape of the WBCs in the image. Unlike in blood, the cells present in bone marrow are packed and often overlapped with each other. The watershed algorithm is used to separate the overlapped white blood cells in the bone marrow slide images. Besides area, several other geometrical features were also extracted from the WBCs include circularity, radius and perimeter. Other color features include mean, standard deviation and variance were also extracted for red, green and blue color respectively. These features were used as input data to the MLP network (standard and hierarchical MLP) and *k*-NN to be classified as Normal, Abnormal type M3 and Other Abnormal bone marrow slide images. Levenberg Marquardt (LM) and Bayesian Regularization (BR) training algorithms were used to train the MLP networks. Standard MLP, network (1A)-BR has managed to achieve the highest accuracy, which is 93.9% on testing dataset in classification of bone marrow slide images into Normal, Abnormal (M3) and Other Abnormal, outperformed MLP network (1A)-LM and *k*-NN classifier (2A). Hierarchical classifier, MLP network (1B) and (1C) has managed to achieve an average accuracy of 99.8% on training and 98.57% in testing outperformed the *k*-NN (2B) and (2C). In general, MLP networks have outperformed the KNN classifiers in the classification tasks of the bone marrow slide images.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Cancer is a disease that conjures up deep fears of a silent killer that creeps up on us without warning. Cancer is a leading cause of death worldwide and accounted for 7.6 million deaths in 2008 (World Health Organization, 2011). Nearly 70,000 new cancer cases were diagnosed among Malaysians in Peninsular Malaysia between 2003 and 2005, according to a report released in early 2008 on the incidence of the disease in West Malaysia (National Cancer Council, Malaysia, 2011). Lung cancer, breast cancer, leukemia and cervical cancer are the most widespread cancers that occur in Malaysia.

Leukemia is a type of blood disease or so-called cancer of the blood. It is the most common form of childhood cancer. In Malaysia, a total of 529 cases of myeloblastic leukemia and 433 cases of lymphocytic leukemia were reported comprising 4.5% of the total number of cancers (National Cancer Registry, Malaysia, 2003). Males predominated over females at a ratio of 1.7:1 for lymphocytic leukemia and 1.1:1 for myeloblastic leukemia. It was the 4th most common cancer with 7.1% in males after lung cancer, 13.8%, nasopharynx cancer, 8.8% and colon cancer, 7.6%. In females, it was the 7th most common cancer with 4.0% after breast cancer, 31%, cervix uteri, 12.9%, colon cancer, 6%, corpus uteri, 4.3%, rectum cancer 4.1% and ovary cancer, 4.1%.

Generally, there are 4 main types of leukemia; acute lymphoblastic leukemia (ALL), acute myeloblastic leukemia (AML), chronic lymphoblastic leukemia (CLL)