

**DEVELOPMENT OF AN AUTOMATED
INTELLIGENT DIAGNOSTIC SYSTEM FOR
TUBERCULOSIS DETECTION BASED ON
SPUTUM SPECIMEN**

RAFIKHA ALIANA A. RAOF

UNIVERSITI MALAYSIA PERLIS

2014

© This item is protected by original copyright



UniMAP

**DEVELOPMENT OF AN AUTOMATED
INTELLIGENT DIAGNOSTIC SYSTEM FOR
TUBERCULOSIS DETECTION BASED ON
SPUTUM SPECIMEN**

by

**RAFIKHA ALIANA A. RAOF
(0640210101)**

A thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy

**School of Computer and Communication Engineering
UNIVERSITI MALAYSIA PERLIS**

2014

UNIVERSITI MALAYSIA PERLIS

DECLARATION OF THESIS

Author's full name : RAFIKHA ALIANA A. RAOF
Date of birth : 17 NOVEMBER 1979
Title : DEVELOPMENT OF AN AUTOMATED INTELLIGENT DIAGNOSTIC
SYSTEM FOR TUBERCULOSIS DETECTION BASED ON SPUTUM
SPECIMEN
Academic Session : 2006-2014

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

791117-06-5304

(NEW IC NO. / PASSPORT NO.)

Date : _____

SIGNATURE OF SUPERVISOR

PROF. DR. MOHD. YUSOFF B. MASHOR

NAME OF SUPERVISOR

Date : _____

ACKNOWLEDGEMENT

ALHAMDULILLAH, all praises be to Allah, Most Gracious and Most Merciful for giving me the strength and guidance in finalizing this PhD thesis. First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Dr. Mohd. Yusoff bin Mashor for his continuous insight, knowledge, encouragement, guidance and motivation. It is a privilege and a great pleasure to have him as the supervisor. I am grateful to my co-supervisors, namely Professor Dr. R. Badlishah Ahmad and Dr. Siti Suraiya Md. Noor, who have assisted me most generously in the completion of this research. My sincere thanks also go to the staff in Department of Medical Microbiology and Parasitology, School of Medical Science, USM for the assistance. I am also grateful to the Universiti Malaysia Perlis (UniMAP) and Ministry of Higher Education Malaysia (MOHE), for their financial support during the past years.

Special thanks to my beloved husband, Muhamad Mizar Morat who has been my best friend and my true companion. I thank him for his unconditional love, perseverance and never ending support. To my children, Muhammad Rafiq Ezzat, Nur Ezzah Raihana and Nur Effah Raihana who have been my source of joy, keeping my life filled with laughter and love. Dedicated to my parents and in laws for their innumerable prayers and encouragement. They have stood by me in everything I have done, providing constant love and support. My sincere thanks also go to all the members of my research group in the UniMAP Autonomous and Machine Vision Cluster for the teamwork and friendship during hard times and sunny days. *Jazakallah Khairan Kathira.*

TABLE OF CONTENT

	PAGE
THESIS DECLARATION	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xv
LIST OF SYMBOLS	xvii
ABSTRAK	xx
ABSTRACT	xxi
CHAPTER 1 INTRODUCTION	
1.1 Introduction	1
1.2 Tuberculosis Diagnosis System	3
1.3 Problem Statements	5
1.4 Objective of Research	7
1.5 Scope of Research	7
1.6 Thesis Outline	8

CHAPTER 2 LITERATURE REVIEW

2.1	Introduction	10
2.2	Tuberculosis	10
2.2.1	Types of Tuberculosis	12
2.2.1.1	Pulmonary Tuberculosis (PTB)	12
2.2.1.2	Extrapulmonary Tuberculosis (EPTB)	13
2.2.2	Causes, Signs and Symptoms of Tuberculosis	14
2.2.3	Diagnosis and Classification	16
2.2.3.1	Procedures in Diagnosing Pulmonary Tuberculosis	17
2.2.3.2	Tuberculosis Slide Reporting	19
2.3	Digital Image Processing	22
2.3.1	Image Enhancement	22
2.3.1.1	Partial Contrast	24
2.3.1.2	Dark Stretching	26
2.3.1.3	Bright Stretching	27
2.3.1.4	White Balance	28
2.3.2	Image Segmentation	29
2.3.2.1	Colour Thresholding	30
2.3.2.2	Clustering	33
2.3.3	Noise Filtering	35
2.3.3.1	Median Filter	36
2.3.3.2	Mean Filter	38

2.3.3.3	Region Growing	38
2.3.4	Morphological Features for TB Bacilli	40
2.4	Artificial Neural Networks (ANN)	42
2.4.1	Multilayer Perceptron	44
2.4.1.1	MLP Architecture	45
2.4.1.2	Levenberg-Marquadt Training Algorithm	46
2.4.2	Hybrid Multilayer Perceptron (HMLP)	49
2.4.2.1	HMLP Architecture	50
2.4.2.2	Modified Recursive Prediction Error Training Algorithm	51
2.4.3	Extreme Learning Machine	55
2.4.3.1	SLFN Architecture	55
2.4.3.2	ELM Algorithm	56
2.5	Diagnosis System for Tuberculosis detection	58
2.6	Summary	65
CHAPTER 3 IMAGE PROCESSING PROCEDURES FOR ZN SPUTUM SLIDE IMAGES		
3.1	Introduction	66
3.2	Development of Image Processing Procedure	67
3.3	Image Acquisition of Sputum Slide Images	67
3.3.1	Image Capturing Setup	69
3.3.2	Image Properties	72
3.4	Image Processing Procedure for TB Slide Images	75

3.5	Image Enhancement for TB Slide Images	78
3.5.1	White Balance	78
3.5.2	Partial Contrast	85
3.6	Image Segmentation for TB Slide Images	94
3.6.1	Fixed-value Colour Thresholding	95
3.6.2	<i>k</i> -means Clustering	106
3.6.3	Comparison of Image Segmentation Techniques	114
3.7	Noise Filtering	116
3.7.1	Modified Median Filter	118
3.7.2	Automated Seed Based Region Growing	125
3.7.3	Comparison of Noise Filtering Techniques	130
3.8	Conclusion	132
CHAPTER 4 INTELLIGENT CLASSIFICATION PROCEDURES FOR DETECTING TB BACILLI IN ZIEHL-NEELSEN SPUTUM SLIDE IMAGES		
4.1	Introduction	135
4.2	Development of Intelligent Diagnostic Procedure	136
4.3	Feature Extraction for TB Bacilli in Sputum Slide Images	136
4.3.1	TB Bacilli Features	138
4.3.1.1	Size-based Feature Extraction	138
4.3.1.2	Colour-based Feature Extraction	140
4.3.1.3	Shape-based Feature Extraction	143
4.4	Classification of TB Bacilli using Artificial Neural Network	148
4.4.1	Data Preparation as Input to Artificial Neural Network	149

4.4.2	Methodology of Determining the Suitable ANN Structure	150
4.4.2.1	Multi Layered Perceptron Network Trained with LM Learning Algorithm	151
4.4.2.2	Hybrid Multi Layered Perceptron Network Trained with MRPE Algorithm	153
4.4.2.3	Single Layer Feed Forward Neural Network Trained with ELM Algorithm	155
4.4.2.4	Performance Comparison of Various ANN Classifications	157
4.4.3	Feature Selection using HMMLP network for TB Bacilli Classification	158
4.4.3.1	Analysis of Features from Size Category	160
4.4.3.2	Analysis of Features from Colour Category	161
4.4.3.3	Analysis of Features from Shape Category	163
4.4.3.4	Analysis of Combined Selected Features for All Categories	165
4.5	Conclusion	167

CHAPTER 5 AUTOMATED INTELLIGENT DIAGNOSTIC SYSTEM FOR TUBERCULOSIS

5.1	Introduction	169
5.2	Overall Design of Automated Intelligent Diagnosis System for Tuberculosis	170

5.3	Automated Diagnosis System for Tuberculosis based on ZN Sputum Specimen	173
5.3.1	Computer-Assisted ZN Sputum Slide Image Manual Diagnosis	176
5.3.2	Automated ZN Sputum Slide Image Diagnosis	179
5.3.2.1	TB Bacilli Classification Procedure for ZN Sputum Slide Images	183
5.3.2.2	TB Slide Classification Result	185
5.4	Conclusion	190
CHAPTER 6 CONCLUSION AND FUTURE WORK RECOMMENDATION		
6.1	Conclusion	192
6.2	Research Contribution	193
6.3	Future Work Recommendation	195
REFERENCES		196
APPENDIX A		207
APPENDIX B		219
LIST OF PUBLICATIONS		225
LIST OF AWARDS		228

LIST OF TABLES

NO.		PAGE
2.1	Reporting scale for TB diagnosis	20
2.2	Summary of previous work on Screening System for TB Detection	61
3.1	RGB information for images in normal category	96
3.2	RGB information for images in dark category	97
3.3	RGB information for images in bright category	98
3.4	The definition of TP, TN, FP and FN for TB bacilli segmentation	115
3.5	Comparison of overall segmentation performance for two different methods	116
3.6	Comparison of overall noise filtering performance for the different procedures	131
4.1	List of Proposed Features based on its Category	137
4.2	Dataset for training, validation and testing samples	149
4.3	Structure and Parameters used in training phase of MLP network	151
4.4	Result of best structure for MLP network tested in 5 states	152
4.5	Structure and Parameters used in training phase of HMLP network	154
4.6	Result of best structure for HMLP network tested in 5 states	155
4.7	Structure and Parameters used in ELM algorithm	156
4.8	Result of best structure for SLFNN network tested in 5 states	156
4.9	Result of best structure for each ANN	157

4.10	Highest Recognition Performance for Features in Size Category	160
4.11	Highest Recognition Performance for Features in Colour Category	161
4.12	Highest Recognition Performance for Features in Shape Category	163
4.13	Feature Combination Results for Three Categories	166
5.1	TB Slide Classification Performance using Automated Diagnosis System	188
5.2	Sputum Slide Classification Result	188

© This item is protected by original copyright

LIST OF FIGURES

NO.		PAGE
1.1	Estimated TB incident rates in 2011	2
2.1	Main symptoms of PTB	13
2.2	Main sites of EPTB	14
2.3	Droplets nuclei can remain in the air for a long period	15
2.4	Smear microscopy based on Ziehl-Neelsen technique	19
2.5	Partial contrast process	25
2.6	Dark stretching process	26
2.7	Bright stretching process	27
2.8	Thresholding for a single threshold	31
2.9	Thresholding with a pair of threshold	32
2.10	Square mask of $n \times n$ pixels	37
2.11	Growing seed pixel based on eight-surrounding neighbour	39
2.12	Mycobacterium tuberculosis scanned under electron micrograph	40
2.13	An MLP network	45
2.14	Hybrid Multilayered Perceptron Network	50
3.1	Flowchart of development of automated diagnosis system for tuberculosis detection	68
3.2	A hardware set of Leica microscope, digital camera and personal computer interfaced together to acquire the digital images	69
3.3	Samples of ZN sputum slides specimen	70
3.4	Capturing Flow for Smear Area on ZN Sputum Slide Specimen	70

3.5	Original sputum slide images consisting of TB bacilli	73
3.6	TB bacilli and Sputum cells	73
3.7	Negative slide images without TB bacilli	74
3.8	Sputum Slide Image and Its Histogram	75
3.9	The proposed image processing steps for tuberculosis sputum slide images	76
3.10	Samples of original images with their intensity histogram	77
3.11	Samples of Dark Image	79
3.12	Result of applying white balance technique	82
3.13	Result of applying unsuitable percentage values	83
3.14	Result of applying white balance technique on normal image	84
3.15	Result of applying white balance technique on overexposed image	84
3.16	Partial Contrast Technique applied to sputum slide image	89
3.17	Original Underexposed Image and Its Partial Contrast Image	91
3.18	Partial Contrast Technique applied to red component	92
3.19	Result of Partial Contrast	93
3.20	Result of Partial Contrast on Bright Image	94
3.21	Original and reference images from normal, dark and bright category with their respective result after applying rules for normal image	101
3.22	Original dark images with result after applying rules for dark image	102
3.23	Original images from bright category with the result after applying rules for bright image	103
3.24	Segmentation results of enhanced images	104

3.25	Method to retrieve <i>max_blue</i> value from blue component histogram	110
3.26	Original and reference images from normal category with the result after applying set of clustering procedure	112
3.27	Original and reference images from dark category with the result after applying set of clustering procedure	112
3.28	Original and reference images from bright category with the result after applying set of clustering procedure	113
3.29	Output of thresholding stage	117
3.30	Square mask of $n \times n$ pixels	118
3.31	Result of median filter	120
3.32	Result of median filter with different n values	120
3.33	Normal image and its result	123
3.34	Bright image and its result	123
3.35	Dark image and its result	124
3.36	Graph showing the processing time of original and modified algorithm	125
3.37	Growing seed pixel based on surrounding neighbours	126
3.38	Normal image and its result	128
3.39	Bright image and its result	129
3.40	Dark image and its result	129
5.1	Overall hardware connection for the developed automated intelligent diagnosis system for Tuberculosis detection	171
5.2	The automated diagnosis system for tuberculosis detection	172
5.3	Main user interface for automated diagnosis system for TB	175
5.4	User Interface for Manual Image Capturing	176
5.5	User Interface for Manual Slide Navigation in Section B	177

5.6	ZN sputum slide specimen	178
5.7	User Interface for Single Image Capture in Section C	178
5.8	User Interface for Auto Diagnosis Tab	180
5.9	Image Filename Format in Automatic Capturing	180
5.10	Information of auto diagnosis process	181
5.11	Sample of Result for Auto Diagnosis	182
5.12	Flowchart of the Capturing and Diagnosing Process for Each Slide	184
5.13	Flowchart of the Procedure for Classifying the ZN Sputum Slide Images	186
5.14	Result of Partial Contrast on Suspected Negative Images	189

© This item is protected by original copyright

LIST OF ABBREVIATIONS

AFB	Acid-Fast Bacilli
AIDS	Acquired Immunodeficiency Syndrome
ANN	Artificial Neural Network
API	Application Programming Interface
AWB	Automatic White Balance
CCD	Charge-Coupled Device
DLL	Dynamic Link Library
ELM	Extreme Learning Machine
EPTB	Extrapulmonary Tuberculosis
FN	False Negative
FP	False Positive
FPGA	Field-Programmable Gate Array
FPS	Frame Per Second
GN	Gauss–Newton
GUI	Graphical User Interface
HIV	Human Immunodeficiency Virus
HMLP	Hybrid Multi-Layered Perceptron
HSI	Hue, Saturation, Intensity
IR	Infrared
IUATLD	International Union Against Tuberculosis and Lung Disease
k -NN	k -Nearest Neighbors

LM	Levenberg-Marquardt
MLP	Multilayer Perceptron
MMR	Mass Miniature Radiography
MOH	Ministry of Health
MRPE	Modified Recursive Prediction Error
PTB	Pulmonary Tuberculosis
RBF	Radial Basis Function
RGB	Red, Green, Blue
ROI	Region Of Interest
RPE	Recursive Prediction Error
SBRG	Seed-Based Region Growing
SDK	Software Development Kits
SLFN	Single Layer Feedforward Network
SVM	Support Vector Machine
TB	Tuberculosis
TN	True Negative
TP	True Positive
USB	Universal Serial Bus
USM	Universiti Sains Malaysia
WB	White Balance
WHO	World Health Organization
WP	White Point
ZN	Ziehl-Neelsen

LIST OF SYMBOLS

p_k	colour level of the output pixel
q_k	colour level of the input pixel
max	desired maximum colour level in the output image
min	desired minimum colour level in the output image
f_{max}	maximum colour level in an input image
f_{min}	minimum colour level in an input image
T	threshold value
$f(x, y)$	original pixel value
$g(x, y)$	resulted pixel value
T_1	lower threshold value
T_2	upper threshold value
k	Number of clusters
C_k	k^{th} cluster centre
\mathcal{E}	Euclidean distance
$in(x, y)$	value of the input pixel
n_k	number of pixels belonging to centre C_k
w	neighbourhood centred around location input pixels
$p_0(x, y)$	initial seed location
$\nabla^2 V(x)$	Hessian matrix

$e(x)$	vector of network errors
$J(x)$	Jacobian matrix
μ	Marquardt adjustment parameter
I	identity matrix
w_{ij}^1	weights between input and hidden layer
w_{jk}^2	weights between hidden and output layer
w_{ik}^l	weights between input and output layer
b_j^1	thresholds in hidden nodes
v_i^0	thresholds in inputs
n_i	input nodes
m	number of output nodes
n_h	number of hidden nodes
$F(\bullet)$	activation function
$\varepsilon_k(t)$	prediction error
$y_k(t)$	desired outputs
$\hat{y}_k(t)$	network outputs
$\alpha_m(t)$	Momentum
$\alpha_g(t)$	learning rate
$\psi(t)$	gradient of the one step ahead predicted output
$\lambda(t)$	forgetting factor

λ_0	initial forgetting factor
α	Minimum size of the current region
S_k	Size
P_k	Perimeter
\overline{R}_k	Average Red
\overline{G}_k	Average Green
\overline{B}_k	Average Blue
\overline{Hue}_k	Average Hue
\overline{Sat}_k	Average Saturation
\overline{Int}_k	Average Intensity
C_k	Compactness
ε_k	Eccentricity
μ_{pq}	Central Moments
I_k	Hu Moments

Pembangunan Sistem Diagnosis Pintar Automatik untuk Pengesanan Tuberkulosis Berasaskan Sampel Kahak

ABSTRAK

Tuberkulosis (TB) adalah penyakit berjangkit yang amat merbahaya. Diagnosis untuk penyakit TB biasanya dilakukan secara manual oleh ahli mikrobiologi melalui pemeriksaan mikroskopik spesimen kahak pesakit TB untuk penyakit TB pulmonari. Walau bagaimanapun, amalan tersebut memakan masa dan memerlukan tenaga kerja yang banyak. Oleh yang demikian, ia menyebabkan keletihan, beban kerja berlebihan dan sekaligus menyebabkan penurunan prestasi dalam membuat diagnosis. Kajian ini melibatkan pembangunan sistem diagnosis pintar automatik bagi mengesan TB berdasarkan spesimen kahak Ziehl – Neelsen. Sistem yang dibangunkan ini juga dilengkapi dengan sistem rakaman automatik bagi merakam imej spesimen slaid kahak secara automatik menggunakan kanta 40x. Selain itu, kajian ini juga mencadangkan gabungan teknik pemrosesan imej dengan rangkaian neural buatan dalam mewujudkan satu prosedur baru bagi proses mendiagnosis spesimen kahak Ziehl – Neelsen. Teknik peningkatan imej berdasarkan kaedahimbangan putih dan kontras separa telah dicadangkan. Prosedur yang baru untuk teknik peruasan juga dicadangkan berdasarkan gabungan pengelompokan 'k-means', penapis median bersaiz 3×3 dan algoritma ASBRG. Kajian ini juga melibatkan pengekstrakan ciri, di mana ciri-ciri seperti saiz, warna dan bentuk telah dipilih dalam mengklasifikasikan TB 'bacilli' dengan bantuan rangkaian neural buatan. Kajian ini telah mencadangkan penggunaan rangkaian perseptron berbilang lapisan hibrid (*hybrid multilayered perceptron network, HMLP*) untuk proses pengecaman dan pengelasan bacteria TB. Sistem ini sepatutnya akan dapat mengurangkan masalah yang timbul semasa proses diagnosis penyakit TB seperti ketegangan pada mata ahli teknologi kerana pemerhatian melalui kanta mata mikroskop untuk tempoh masa yang panjang. Pengklasifikasian untuk spesimen slaid kahak bagi diagnosis TB menghasilkan keputusan yang baik dengan ketepatan pengelasan lebih daripada 94%. Kajian yang telah dijalankan ini telah menyediakan platform untuk sistem diagnosis pintar automatik untuk mendiagnosis penyakit TB.

Development of An Automated Intelligent Diagnostic System for Tuberculosis Detection based on Sputum Specimen

ABSTRACT

Tuberculosis (TB) is a highly infectious disease. TB diagnosis is usually done manually by microbiologist through microscopic examination of sputum specimen of TB patients for pulmonary TB diseases. However, this practice is time consuming and labour-intensive. Hence, it results in fatigue and work overload to the microbiologists, thus reduces the diagnostic performance. This research involved in the development of automated intelligent diagnosis system for tuberculosis detection based on Ziehl-Neelsen sputum specimen. The system developed is also equipped with automatic capturing system for capturing sputum slide images automatically using 40X lens. Besides that, this study also suggested the combination of image processing technique with artificial neural network in creating a new procedure for diagnosing process of Ziehl-Neelsen sputum specimen. Image enhancement technique based on white balance and partial contrast method has been proposed. A new procedure for segmentation technique was also proposed based on the combination of k -means clustering, 3×3 median filter and automated seed based region growing algorithm. The study also includes feature extraction where features such as size, colour and shape were chosen in classifying TB bacilli with the aid of artificial neural network. This research proposed to use HMLP network with MRPE algorithm for detection and classification of TB bacilli. The system is supposed to reduce the problems arise during the diagnosis of tuberculosis disease such as avoidance of eye fatigue to the microbiologist due to observing through the microscope eyepiece for a long period of time. It has been shown that the classification for sputum slide specimen for TB diagnosis produces good results with classification accuracy of more than 94%. These findings suggest the potential use of this software in diagnosing pulmonary TB disease. The conducted research has provided the platform for automated intelligent system to diagnose tuberculosis disease.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Tuberculosis (TB) is one of the oldest diseases known to affect humans. At the beginning of the new millennium, despite efforts in the past decade to bring the problem under control, TB remains the most important infectious disease in the world. WHO (2011) mentions in its report that globally, in 2010, there were 8.8 million TB occurrences, with 1.1 million deaths from TB among Human Immunodeficiency Virus (HIV)-negative people. On top of that, HIV-associated TB contributed of an additional 0.35 million deaths. 65% of the estimated number of incident cases in 2010 comes from 5.7 million notifications of new and recurrent cases of TB. TB is affecting mostly young adults in their most productive years. The vast majority of TB deaths are happening in the developing world, with more than half occurring in Asia, thus making TB as a disease of poverty. In 2010, the majority of the estimated number of TB cases occurred in Asia (59%) and Africa (26%). There are also cases found in the Region of Eastern Mediterranean (7%), the Region of Europe (5%) and the America Regions (3%). Five countries with the highest number of occurrence cases in 2010 were India (2.0–2.5 million), China (0.9–1.2 million), South Africa (0.40–0.59 million), Indonesia (0.37–0.54 million) and Pakistan (0.33–0.48 million). An estimated one quarter (26%) of all TB cases worldwide occurred in India along, with another 38% of the cases found in China and India (WHO, 2011). Fig. 1.1 shows the estimated TB incident rate in 2011.