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

RAFIKHA ALIANA A. RAOF

UNIVERSITI MALAYSIA PERLIS

2014

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

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

CHAPTER 2 LITERATURE REVIEW

LIST OF TABLES

LIST OF FIGURES

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
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 pemprosesan imej dengan rangkaian neural buatan dalam mewujudkan satu prosedur baru bagi proses mendiagnosis spesimen kahak Ziehl – Neelsen. Teknik peningkatan imej berdasarkan kaedah imbangan 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 pengkelasan 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 mendiagnos 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.

TB still remains a major global health problem despite the availability of highly efficacious treatment for decades. In 1993, an estimated 7–8 million cases and 1.3–1.6 million deaths occurred each year. This situation forced the World Health Organization (WHO) to pronounce TB as a global public health emergency, (WHO, 2007).

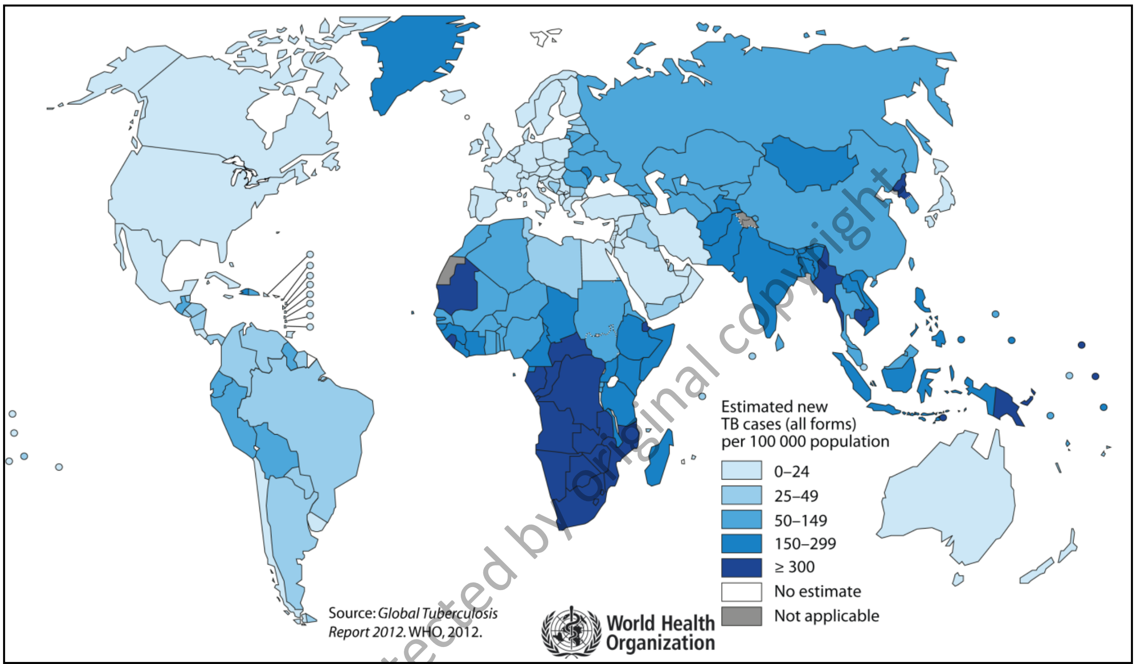

Figure 1.1: Estimated TB incident rates in 2011 (WHO, 2012)

More than one third of the world's total population, which are approximately two billion people, are infected with TB. In their lifetime, one in every 10 of those people will turn out to be sick with active TB. Those who are living with HIV are having a higher risk. In order to achieve the target under the Millennium Development Goals (MDG), WHO is working with other agencies. WHO also aims to reach all patients through primary health care and systems. TB spreads through the air and therefore is contagious. Each person with active TB can infect on average 10 to 15 people a year if it is not treated. There are two TB targets that are set for 2015 and the world is currently on the right schedule to accomplish them. The first target is the MDG that aims to reverse and halt global incidence (in comparison with 1990); and the