BORANG PENGESAHAN STATUS TESIS

JUDUL: THE GENE EXPLORATION OF CHRONIC DISEASES USING SELF-ORGANIZING MAP

SESI PENGAJIAN: 2012/2013

Saya NURLIYANA BINTI MUTY
(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hak milik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
_____	TIDAK TERHAD	

_____ _____

(TANDATANGAN PENULIS) (TANDATANGAN PENYELIA)

Alamat tetap: No.15, Jalan Gapi 1E/5 Antara Gapi, 48200 Serendah,Selangor.

DR.ZURAIDA ABAL ABAS

Nama Penyelia

Tarikh : _____ Tarikh : _____

CATATAN: *Tesis dimaksudkan sebagai Laporan Projek Sarjana Muda(PSM)
**Jika Tesis ini SULIT atau TERHAD, sila Lampirkan surat daripada pihak berkuasa.

© Universiti Teknikal Malaysia Melaka

THE GENE EXPLORATION OF CHRONIC DISEASES USING SELF ORGANIZING MAP

NURLIYANA BINTI MUTY

This report is submitted in partial fulfilment of the requirements for the Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2013

© Universiti Teknikal Malaysia Melaka

DECLARATION

I hereby declare that this project report entitled

THE GENE EXPLORATION OF CHRONIC DISEASES USING SELF ORGANIZING MAP

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date: _____

(NURLIYANA BINTI MUTY)

SUPERVISOR : _____ Date: _____

(DR ZURAIDA ABAL ABAS)

© Universiti Teknikal Malaysia Melaka

DEDICATION

Alhamdulilah and thank you to my family especially my mum Mdm Wan Azimah Binti Wan Muda and my dad Mr Muty Bin Othman. Thank you so much for always stand by me .To my lovely sister Nurul Ain Binti Muty thank you so much.

© Universiti Teknikal Malaysia Melaka

ACKNOWLEDGEMENTS

I would like to thank Dr Zuraida for helping me a lot in my project and finally this work done very well. To my friend Siti Aisyah, you have gave me the strength to work out with this project.

© Universiti Teknikal Malaysia Melaka

ABSTRACT

Coronary heart disease, diabetes, high blood pressure and chronic kidney are the most highly ranked chronic diseases. The diseases can also be represented by the gene. Furthermore, the gene expression is the way of getting information of gene from the protein. This protein will give the information of the diseases occur in the human body. Thus, the method of detecting diseases in human body is the genetic testing. The genetic testing must be proceeding but the probe set must be known. The probe set is the measurement in DNA Microarray. The Self-organizing map will be the technique chosen to find the best probe set to represent a gene. The gene is chosen based on the attributes and description taken to finalize the analysis. The gene code changes because of the order of mutation. Thus, the mutation of gene in human body really gives the picture of disease. Probe used to explain to make a measurement in a gene expression experiment. Probe-set is a package of two or more probes that are used to measure a molecular. The dataset will be preprocessing in term of probe set. The analysis phase gives the best picture in other to determine the visualization of the data. This chapter defines the techniques used in the MATLAB. The Unified identified matrix is used for the visualization of the self-organizing map

© Universiti Teknikal Malaysia Melaka

ABSTRAK

Penyakit jantung, kencing manis, tekanan darah tinggi dan buah pinggang kronik adalah penyakit kronik yang paling sangat kedudukan. Penyakit-penyakit ini juga boleh diwakili oleh gen. Tambahan pula, gen adalah cara untuk mendapatkan maklumat gen daripada protein. Protein ini akan memberikan maklumat daripada penyakit berlaku dalam tubuh manusia. Oleh itu, kaedah mengesan penyakit dalam badan manusia adalah ujian genetik. Ujian genetik mesti meneruskan tetapi set siasatan mesti diketahui. Set siasatan ialah ukuran, dalam DNA Microarray. Peta sendiri menganjurkan akan menjadi teknik yang dipilih untuk mencari yang terbaik siasatan set untuk mewakili gen. Gen ini dipilih berdasarkan ciri-ciri dan penerangan yang diambil untuk menyelesaikan analisis. Kod gen berubah kerana perintah mutasi. Oleh itu, mutasi gen dalam tubuh manusia benar-benar memberi gambaran penyakit. Siasatan yang digunakan untuk menerangkan untuk membuat ukuran dalam eksperimen gen. Siasatan-set adalah pakej dua atau lebih probe yang digunakan untuk mengukur molekul. Dataset akan prapemprosesan dari segi set siasatan. Fasa analisis memberi gambaran yang terbaik di lain-lain untuk menentukan visualisasi data. Bab ini mendefinisikan teknik-teknik yang digunakan dalam MATLAB. Matriks Bersepadu dikenal pasti digunakan untuk visualisasi peta diri penganjur.

© Universiti Teknikal Malaysia Melaka

TABLE OF CONTENT

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

© Universiti Teknikal Malaysia Melaka

LIST OF TABLES

© Universiti Teknikal Malaysia Melaka

LIST OF FIGURES

© Universiti Teknikal Malaysia Melaka

LIST OF ABBREVIATIONS

SOM	-	Self-Organizing Map
CHD	-	Coronary Heart Disease
CKD	-	Chronic Kidney Disease
DM	-	Diabetes Mellitus
HBP	-	High Blood Pressure
NHBI	-	National Heart Blood Institute

© Universiti Teknikal Malaysia Melaka

LIST OF ATTACHMENTS

© Universiti Teknikal Malaysia Melaka

CHAPTER I

INTRODUCTION

1.1 Project Background

Nowadays, the chronic disease is expanding throughout the world. The four current chronic diseases are coronary heart disease, diabetes mellitus, high blood pressure and chronic kidney disease. For this project, the self-organizing map is proposed for this analysis. Therefore, Self- Organizing Map is widely used in this analysis because of the best visualization from high dimensional data to low dimensional data.

Sometimes, there are several person do not have any information about their disease, and finally the person are diagnosed to the related disease. Here, in this research the normally diagnosed disease will undergo the same test, but a part of that there are another level of diagnose the disease through genetic testing. There are clustering methods that can be used in this research it is the unsupervised method that related with the analysis of data process. So in this technique, the clustering is proposed for this research. The highlight for this technique is that the best clustering results will be obtained. Therefore, the category will be classified into certain attributes. On the other hand, Self- Organizing Map is also used in this research because of the eligible used that can transform the high dimensional data to the low dimensional data.

© Universiti Teknikal Malaysia Melaka

1.2 Problems Statement(s)

Currently the approaches normally to diagnose the person with the diseases for example coronary heart disease will undergo for EKG test to diagnose the disease. As well as, the diabetes mellitus will have the fasting glucose test to detect the percentage of glucose in the person body. Furthermore, the high blood pressure will also measure with the cuff and gauge. The person will also undergo the urine test if they have the symptoms as getting chronic kidney disease. But, in this we propose to another level by genetic testing. The genetic testing will identify the chromosome, gene and DNA in the person if any disease infected. In other to have that test, which is the gene expression analysis, we need to identify the best probe set for the test.

1.3 Objective

1) To explore current chronic diseases
2) To determine the best probe set for gene expression analysis by using Self Organizing Map
3) The initial phase for further clinical research for genetic testing

1.4 Scope

The scope for this project is concentrated on the use of modification for the clinical research in the community.

© Universiti Teknikal Malaysia Melaka

1.5 Project Significance

This project will bring benefits to the community especially for the patient that have several diseases related. By this intelligent system, it can help the computer scientist, clinical researcher and geneticist in other to find the answer instead of typical epidemiology.

1.6 Expected Output

1. The findings will help Malaysia center of research in knowing the diseases detected by genes.
2. This will give the big opportunity between the health knowledge together with the data analysis.
3. It would give the new model in defining related diseases.
4. The computer scientist, geneticist and clinical researcher can work together in helping the medical diagnosed of disease.

1. Project Publications

- Ministry Of Health Malaysia

2. Specific or Potential Applications (if any)

- The result produced by this research is beneficial to the ministry of health especially to the hospital for the department of clinical research in Malaysia this findings can enhance the development of epidemiology in our country.Besides, it can be implemented in other domains such as dietary pattern for analysis.

© Universiti Teknikal Malaysia Melaka

3. ***Impact on Society,Economy and Nation***

- It is believed that this findings will give a positive impact for the clinical research in making better decisions.This will also open up the malaysian eyes to take care of their health and lead good lifestyle.

1.7 Conclusion

The aim of this research is to prove how useful the method of clustering in finding the best solution for the health management. It will be useful in determine the specific disease in other to get the same relationship with the use of data mining.

© Universiti Teknikal Malaysia Melaka

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

It is common knowledge that the disease has become over the last few decades, it is one of the most important Malaysian health problem. It is the silent killer that affected among the Malaysian. This is the mainly the result that this explosive increase result that the number of people diagnosed with the disease is definitely increasingly by few decades. As a matter of facts, it is the developing countries that present themselves with the highest rate of the disease. In this research the review of four diseases is a must in finding the large knowledge about it

2.2 The Diseases

The disease will be identified with the history, the complication, and the treatment for the disease. The diseases are coronary heart disease, diabetes mellitus, high blood pressure and chronic kidney disease. The chronic disease such as heart coronary, high blood pressure, hypertension and chronic kidney disease is the tumor that causes many dead in over the world. It is necessary for the people to have knowledge about this unpredictable disease. However, to do this, we will explore the use of data on coronary heart, diabetes, high blood pressure and chronic kidney disease. The tool of Mat lab is use in other to show the representation for the result analysis. The self- organizing map will be involved in the data analysis.

© Universiti Teknikal Malaysia Melaka

2.2.1 Coronary Heart Disease

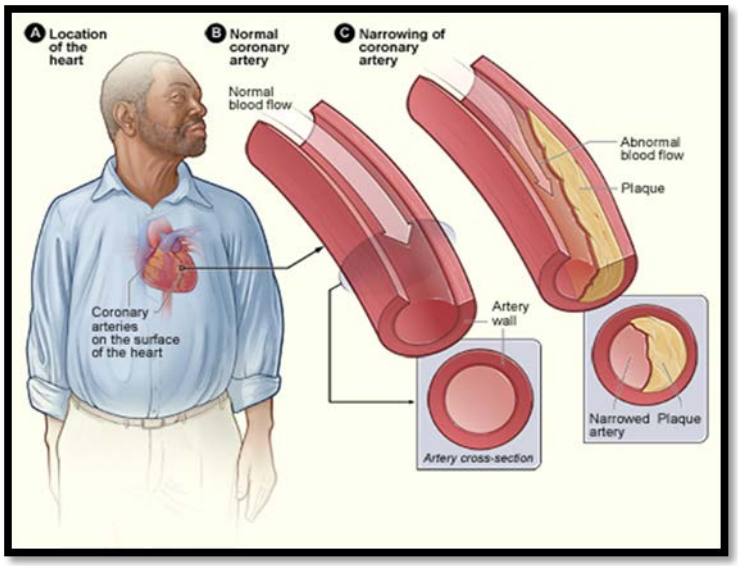

Figure 1: The Structure of artery (Hage et al. 2009)

The coronary heart disease has become one of the highest ranking diseases among Malaysian. It is surprisingly every person is diagnosed with this heart coronary. The coronary heart disease (CHD) happened because of the narrowing or blockage of the coronary arteries, it is caused by the atherosclerosis(Flaherty et al. 2009). Thus, the figure from National Heart, Lung and Blood Institute shows the occurrences of the heart coronary disease. From the figure 1 of National Heart Lung Blood Institute of United States, The figure A shows how the location of the heart in the body. Then, for the figure B show the normal artery with usual blood flow. While in the figure C, show how the artery being affected by the plaque that causes the miserable for heart to pump in the whole body.

From the perspective of the research, the coronary heart disease occurs because the damage happened in the arteries. It is because of the bad lifestyle habits. Sometimes a person who is smoking can become one of the causes. Thus smoking can clog the blood vessels and caused to inefficient cholesterol ranges, and this will high the blood

© Universiti Teknikal Malaysia Melaka

pressure. Thus, it can prevent the oxygen to reach the artery. Thus, women also include in the diagnosis patient having coronary heart disease. It happened because of the occurrences in the body that cannot work in a normal condition. The basic feature of the coronary heart disease is the pain in the chest, this normal sign of the person is infected by the disease. It is because of uncomfortable feeling the chest that likely occurs in few minutes. Thus, the people also feel the shortness of the breath. It is because the heart cannot pump enough blood together with the body needs.

There are two stages, the first stage consists of the medical and family histories, and then the doctor will ask about the risk factors for coronary heart disease, the physical exam and also the tests and procedures taken to identify the disease(Anon n.d.). Thus, for the second stage there will be stage which is first the EKG (Electrocardiogram). It defines the hearts electrical activity.

This show the heart is beat. Furthermore, it will detect the flow of the heart based on electrical signs. Second, the stress testing it is basically about the exercise to ask the person heart to work hard and beat very currently while the test is done(Shiba and Shimokawa 2011). Then, if the person cannot do the exercise, the medicine will be given in other to raise the heart rate. Thus, the stress tests show how the inefficiency in the heart levels. The Electro-beam computed tomography. The test looks on the particles of calcium based in the coronary arteries. The calcium is mutual sign of CHD.

© Universiti Teknikal Malaysia Melaka

2.2.2 Diabetes Mellitus or Diabetes

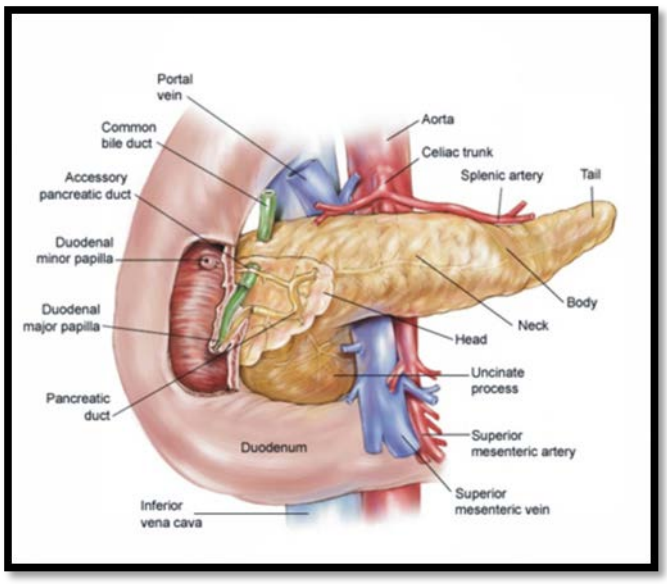

Figure 2: The structure in pancreas (Howard et al. 2010)

The Diabetes has been spread as one of the killer in the people all over the world. This will also include our country, Malaysia among the other developed countries. The aspect of complication and treatment will take place in the research.. But the awareness among people is not efficient. The homeostasis glucose is the measurement the level of blood glucose in a normal level will be normal for the whole day after eating(Mohieldein, Alzohairy, and Hasan 2011). There are about three process in a human body, if this three process be effected this will caused the diabetes. From the figure 2, the health info shows that the occurrences of diabetes.

Three process:

1) Secretions insulin in pancreas gland will stimulate by the food taken

2) This will increase the taken of the glucose by the fat cell, muscles cell and hepatic cell.

3) Inhibition of gluconeogenesis and glycogenesis process and thus minimizes hepatic glucose production

© Universiti Teknikal Malaysia Melaka

The insulin will allow the glucose to release out from the blood. For diabetes, the glucose in the blood cannot regulate in the good order function. From the figure 3, medicine net, it shows the type 1 diabetes (TID). The body is not producing the insulin. It caused during the childhood and adolescence. So the diabetes requires insulin treatment daily to sustain life. This woman will have large babies and this will make them have type 2 diabetes. This happened because of the insulin level is a hormone that conducted by pancreas in order to control the blood glucose. It may be reveal as a little insulin resistance. They will have excess thirst, the person tend to drink a lot of water. They will get tiredness. The person is hunger on the foods that they are taken. On the other hand, they will also having urinate different from usual.

They will also experience the weight having loss increasingly. This fasting glucose test is taken for two ways, there are after the person is not eat anything at least about 8 hours usually the fasting moments and maybe randomly selected time.

Then this test will measure how the blood test from the fasting blood glucose level, the level around 70 and 100 milligrams (mg/dl) is the basic level. So if the person not in the range of this normal level that person will be diagnosed as a diabetes, they are pre diabetes from the level of 100-125 and the 126 higher show that the person is having diabetes(Awah, Unwin, and Phillimore 2009).

The sample from the blood test is taken by the finger stick , this will measure how the diabetes level in the body. There are the measurement of the HbA1c indicates the 5.6% or less is normal level. Then, for the pre diabetes the level is between 5.7% and 6.4 %. While, for the person who is diagnosed to have diabetes, they will have the level of 6.5% or higher this be sure that the person is definitely having diabetes.

© Universiti Teknikal Malaysia Melaka

The oral glucose test (OGTT). This test is taken based on the blood test of the person will be given the glucose to drink. Then, after 30 to 60 minutes will measure how the blood level shows the 75 gram is the normal. While for the range between 140-200 mg/dl is the level for pre diabetes. Then, the person is diagnosing to be the diabetes, if it is 200 mg/dl higher.

2.2.3 High Blood Pressure

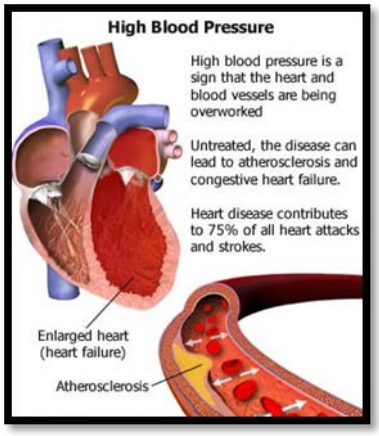

Figure 3 : The artery of coronary(Olafiranye et al. 2011)

Hypertension or high blood pressure is the conditions where the tense of blood pertain through the walls of the blood arteries as the heart will pumps the blood. Furthermore, it will affect the whole body. The figure 3 from the herbal medicine shows how the blood pressure signs in the heart.

Basically the blood pressure is depending on the patient situation. It will get lowers while the person is sleep and rise after wake up. There is also the pre-hypertension which is the person show the abnormal level of the high blood pressure. It is also depend on the person age and family history. Thus, it is also because of the high level of salt taken.

© Universiti Teknikal Malaysia Melaka