Evaluating the Development of Diabetes Risk Using Probability Technique

Noorlin Mohd Ali, Siti Zanariah Satari, Noorhuzaimi@Karimah Mohd Noor & Lee Hooi Yeoh
Faculty of Computer Systems & Software Engineering
Universiti Malaysia Pahang
Locked Bag 12
25000 Kuantan Pahang
MALAYSIA
noorlin@ump.edu.my; zanariah@ump.edu.my; nhuzaimi@ump.edu.my; hooichie@yahoo.com

Abstract - The aim of this study is to develop a prototype in determining the development of diabetes risk using probability technique. Diabetes has becomes one of the major health problem in Malaysia. An estimated 100 million people are affected by diabetes mellitus world wide of whom 7 million are in the ASEAN region. In Malaysia, the prevalence is about 4%. A significant observation is that prevalence (and incidence) of this chronic disease is on the increase. A similar trend is observed in this country. In 1960 prevalence was about 0.65%; in 1982, 2.1% while a limited population-based study in 1984 showed a prevalence of nearly 4% [1]. The main factors that develop diabetics are age, BMI, glucose level, blood pressure and cholesterol level.

In order to overcome the problem, a system that able to determine the development of diabetes risk is needed. The actual system acts as the awareness tool to help user determines the risk of diabetes development in their body. In this study, the prototype has applied knowledge based expert system technique. In order to derive the risk decision, the probability technique has been implemented by using interpretation values on each of the rules set. The range of the risk has been classified by using quartile and percentile technique. Based on the study, the decision risk on diabetics' development is classified by *low risk*, *medium risk*, *high risk* and *very high risk*. The detail outcome and the result from the experiments are discussed in this paper.

Keywords
Diabetes, probability, statistical, quartile and percentile technique, knowledge based.

INTRODUCTION

Diabetes is a disease in which blood glucose levels are above normal. Most of the food we eat is turned into glucose, or sugar, for our bodies to use for energy. The disease becomes one of the major health problems nowadays. For certain extends, it can cause serious health complications such as heart disease, blindness, kidney failure, and lower-extremity amputations.

There are many risk factors that develop diabetes; obesity, increased age, high blood pressure and high cholesterol, and high blood glucose level.

Table 1.0 The risk factor of diabetes development [2]

Risk factor	How does it develop to diabetes
Obesity	This is the number one risk factor for diabetes. More weight means increased insulin resistance, because fat cells are not as receptive to insulin as muscle cells are.
Increased Age	The older we get, the greater our risk of diabetes. Scientists theorize that the pancreas ages right along with us, and doesn't pump insulin as efficiently as it did when we were younger. Also, as our cells age, they become more resistant to insulin.
High blood pressure and high cholesterol	These are the hallmark risk factors for many diseases and conditions, including diabetes, cardiovascular conditions such as heart disease and stroke risk.
High blood glucose level	The level of blood sugar control predicts the onset and severity of diabetes-related complications for both types of diabetes.

The former method in determining the disease is strongly depends on the physician's knowledge and experience [6] [8]. The determining task becomes difficult since a lot of factors should be considered. The physician needs to analyze and compare the previous patient record with the current record in order to determine the risk of the disease.

One of the major opportunities in the community is the explosion of data. In order to appreciate the data existence, data mining and knowledge discovery techniques have been applied to the system development.

The implementations of expert system in diagnosing medical problems are increasingly being used. Expert System is the shell of knowledge based system. It is a computer program that contains some of the subject-specific knowledge of one or more human experts [3]. The most common form of rule based system is a program made up of a set of rules that analyze information (usually supplied by the user of the system) about a specific class of problems. A system that uses human knowledge captured in a computer to solve problems that ordinarily require human expertise. It is a system that utilizes reasoning capabilities to reach conclusions [3].

Expert system consists of three principal parts: the knowledge base and the inference engine and user interface. The knowledge base of expert systems contains both factual and heuristic knowledge [3]. Factual knowledge is that knowledge of the task domain that is widely shared, typically found in textbooks or journals, and commonly agreed upon by those knowledgeable in the particular field [4].Heuristic knowledge is the less rigorous, more experiential, more judgmental knowledge of performance. In contrast to factual knowledge, heuristic knowledge is rarely discussed, and is largely individualistic. Knowledge representation formalizes and organizes the knowledge. One widely used representation is the production rule, or simply rule. A rule consists of an IF part and a THEN part (also called a condition and an action). The IF part lists a set of conditions in some logical combination. The piece of knowledge represented by the production rule is relevant to the line of reasoning being developed if the IF part of the rule is satisfied; consequently, the THEN part can be concluded, or its problem-solving action taken. Expert systems whose knowledge is represented in rule form are called rule-based systems [4].

In this study, the prototype of knowledge based expert system that determines the risk of diabetes is developed. A number of 200 data have been collected in two areas. In developing the prototype, the uncertain conclusion is established by using probability technique, while the range of data are classified by using quartile and percentile technique.

BODY OF PAPER

In developing the rules, data have been collected in two states; Kajang, Selangor and Kota Kinabalu, Sabah.

Table 2.0 Age factor for diabetic in two different areas

Age	Selangor		Sabah	
	Total	Percentage (%)	Total	Percentage (%)
<30	3	2.0	10	5
30-39	36	18	28	14
40-49	45	23	44	22
50-59	66	32	65	32
>=60	50	25	53	27
Total	200	100	200	100

Table 2.1 BMI for the diabetic in two different areas

BMI (kg/m2)	Selangor		Sabah	
	Total	(%)	Total	(%)
<18.5	30	15	15	8
18.5-24.9	33	17	22	11
25-29.9	54	27	66	33
>=30	83	41	97	48
Total	200	100	200	100

Table 2.2 Glucose Level after meal for the diabetic in two different areas

Glucose Level (mm/gl)	Selangor		Sabah	
	Total	(%)	Total	(%)
4.4-6.6	42	21	28	14
6.7-10.0	112	56	101	50
>10.0	46	23	71	36
Total	200	100	200	100

Table 2.3 Blood Pressure Level for the diabetic in two different areas

Blood Pressure(mmHg) Systolic /Diastolic	Selangor		Sabah	
	Total	(%)	Total	(%)
119-129/79-84	75	37	59	30
130-139/85-89	56	28	66	33
140-159/90- 99(or higher)	69	35	75	37
Total	200	100	200	100

Table 2.4 Cholesterol Level for the diabetic in two different areas

Total Cholesterol Level (mg/dl)	Selangor		Sabah	
	Total	(%)	Total	(%)
<200	**23**	**15**	**35**	**18**
200-249	**78**	**38**	**88**	**43**
>249	**99**	**47**	**77**	**39**
Total	**150**	**100**	**150**	**100**

Each of the disease factors is combined in order to derive the risk of diabetes development. While dealing with uncertainty information in developing the rules, the probability technique has been applied.

Probability is a quantitative measure of how likely the event is to occur [5]. It compares the number of ways an event can happen to the number of possible outcomes. It allows to quantifying the strength and or confidence in the conclusion.

$$\text{Probability of the event (P(A))} = \frac{(n(A))}{(n(S))} \quad (1)$$

where (n(A)) is the number of favorable outcomes, and (n(S)) is number of the total outcomes.

The probability value for each factor is shown in the following tables.

Table 2.5 The probability value of age factor

Diabetes Risk Factor Age	Probability
<30	0.03
30-39	0.16
40-49	0.22
50-59	0.33
>=60	0.26

Table 2.6 The probability value of BMI factor

Diabetes Risk Factor BMI(kg/m2)	Probability
<18.5	0.11
18.5-24.9	0.14
25-29.9	0.3
>30	0.45

Table 2.7 The probability value of glucose level factor

Diabetes Risk Factor Glucose Level(mm/gl)	Probability
4.4-6.6	0.18
6.7-10.0	0.53
>10.0	0.29

Table 2.8 The probability value of blood glucose factor

Diabetes Risk Factor Blood Glucose(mm/Hg)	Probability
119-129/79-84	0.34
130-139/85-89	0.31
149-159/90-99	0.36

Table 2.9 The probability value of cholesterol level factor

Diabetes Risk Factor Cholesterol Level	Probability
<200	0.15
200-249	0.42
>249	0.44

The sample of the developed rule is shown in the following figure:

IF age is <30
AND BMI <18.5
AND glucose level 4.4-6.7
AND blood pressure is 120-129
AND cholesterol is <200
THEN you are at 0.4 risks to develop diabetes.

Fig. 1 Example System Rule

In determining the risk result, the statistical technique of quartile and percentile has been applied. The range of the data has been shown in the Table 2.9.

$$Q1 = \frac{Xn}{4} \quad (2)$$

where Q1is the ¼ quartile from the total of data;

$$Q2 = \frac{X\,2n}{4} \quad (3)$$

where Q2 is the ½ quartile from the total of data;

$$Q3= \frac{X\,3n}{4} \qquad (4)$$

where Q3 is the ¾ quartile from the total of data.

Table 2.9 Range of the result based of quartile of data

Range	Result
<Q1	Low Risk
Q1	Medium Risk
Q2	High Risk
Q3	Very High

CONCLUSION

In this study, the rule based expert system technique has been applied in determining the risk of diabetes development. The knowledge on the disease is gained from the expertise and the data has been collected in two areas; in Kajang Selangor and Kota Kinabalu Sabah. In drawing the rules into conclusion, each of the factors that contribute to the diabetes development is set with the uncertainty value by using probability technique. The range of the risk result has been classified as *low risk*, *medium risk*, *high risk* and *very high risk* by using quartile and percentile technique.

On further data exploration, it is hope that other statistical and artificial intelligence technique could be applied.

REFERENCES

[1] Mustaffa Embong, Diabetes in Malaysia – Problems and Challenges, 2002, http://www.mma.org.my/mjm/1_diabetes_90.htm

[2] Word Health Organization: Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications Part I : Diagnosis and Classification of Diabetes Mellitus. 1999, http://www.staff.ncl.ac.uk/philip.home/who_dmc.htm

[3] Durkin, J., Expert System: Design and Development, New Jersey:Prentice Hall, 1994.

[4] Harmon, P. and King, D., Expert System:Artificial Intelligence in Business, John Wiley & Sons, Inc. New York, 1985.

[5] Myers Walpole, Probability and Statistics for Engineering and Scientist, Pearson Prentice Hall, 1993.

[6] Kemal P. and Salih G., "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease", Digital Signal Processing, 17 (4), (July 2007), pp. 702-710.

[7] M.N. Kamel Boulos, F.E. Harvey, A.V. Roudsari, R. Bellazzi, M.E. Hernando, T. Deutsch, D.G. Cramp and E.R. Carson, "A proposed semantic framework for diabetes education content management, customisation and delivery within the M2DM project", Computer Methods and Programs in Biomedicine, 83 (3), (September 2006), pp. 188-197.

[8] Kemal P., Salih G., and Ahmet A., "A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine", Expert Systems with Applications, 34 (1), (January 2008), pp. 482-487.