Prediction System Kidney Complication of Diabetes Patients from Big Data with Data Visualization Technique

Wasun Saithong[1]
and Sooksawaddee Nattawuttisit[2]
Faculty of Information Technology,
Sripatum University, Bangkok, Thailand
[1]wasun.sai@spulive.net
[2]sooksawaddee.na@spu.ac.th

Abstract - **This research aimed to predict the possible results of kidney complication analysis in diabetic patients by using data visualization technique. The total sample size was 6 million records, and it consisted of information diabetes mellitus (DM) gathered from 43 File of health service units nationwide where it was stored in medical and health data warehouses during the fiscal year 2013–2018. During the experiment, the researcher could visibly investigate the complicated factors such as data structures, the completeness, and the accuracy of diabetes mellitus are linked to the trend of complications of diabetes, including patients' gender. However, the relationship between other factors, such as the amount of information and time might not be associated with diabetes in statistical terms. Therefore, data visualization is one of the propitious techniques could support the analysts to examine the gigantic datasets of complications that might occur with diabetic patients. Finally, it could reduce the number of illness and death by disease and complications that would lead a way to solve some of the health problems in the future.**

Keywords - **Prediction, Kidney, Complication, Bigdata, Visualization**

I. INTRODUCTION

Diabetes is a chronic disease that is a major public health problem in the world. Is a rapidly spreading threat to the world. Impact on economic development greatly. The International Diabetes Federation (IDF) reports that 4 million people die every year from diabetes every year, an average of 8 seconds per person. More than 300 million people are diagnosed with diabetes. People in low- and middle-income countries have a higher chance of developing diabetes than those in high-income countries 10-20 years of age [1].

From the Bureau of Policy and Strategy the Ministry of Public Health, reporting in 2009, there were 7,019 people died of diabetes, representing 19 cases per day. During the past 10 years (1999-2009), it was found that diabetic patients in hospitals under the Ministry of Public Health increased 4.02 times [2]. In 2009, people with diabetes Received treatment at 558,156 times in hospitals under the Ministry of Public Health, representing a rate of 64 times/hour [3].

From the comparison of the results of the survey of Thai health status (age 15 years and over), the 3rd year (2003-2004) and the 4th time of the year (2008-2009) found that the prevalence rate of diabetes is not the difference is 6.9% and it is estimated that more than 3.46 million people aged 15 years and

older are facing diabetes. And when separated by sex, the prevalence rate in males decreased from 6.4% to 6%. The prevalence rate in females increased from 6.9% to 7.7% when separated by region. And ordered according to the prevalence of diabetes as follows: Bangkok, Central, North-East, North, and South the prevalence rate is 9.2%, 7.6%, 7.0%, 5.7%, and 5.0% respectively [4].

In addition, the results of the Thai health survey (age 15 years and over) found that the third rank is people who do not know that they have diabetes, 56.6% compared to the 4th survey, found that better than diabetics. Didn't know that they had diabetes before 31.2%, or that one in three diabetics didn't know they had diabetes.

The patients who knew that they had diabetes and received glucose-resistant treatment less than 126 mg/dl (less than 126 mg/dl) accounted for 28.5 percent of patients because some people with the disease Diabetes is not known to have diabetes. Causing blood sugar levels to be ignored until having high blood sugar levels for a long time.

The result of complications of the eyes, kidneys, feet, nervous system and brain arteries in diabetic patients Increasing the rate of sickness and death. Based on the data on diabetes screening in the Royal Water Project Public Health Concern (2011) in the population aged 35 years and over, 18,943,581 people found that 1,581,857 people with diabetes and 277,020 people have complications. Accounted for 17.51 percent, with eye complications 23.09 percent, kidneys 25.09 percent, feet 17.13 percent, heart 10.21 percent, brain 10.12 percent, and other organs 21.19 percent.

Diabetes is a non-communicable disease (chronic) that is a problem Important public health of the world Prediction of the potential complications of diabetes from a large database that occurs each year by using visualization techniques to show the problem of distribution and prevalence, will help to plan problems and reduce complications that will occur in patients in the future [5].

II. LITERATURE REVIEW

In theorem, Forecasting definitely refers to predictions in various ways, the tendency of things to happen in the operation of information. The important thing to do. Before planning or preparing to start doing something for accuracy and precision in decision making. Forecasting is about predicting the future. Prophecy and truth may be possible. However, good predictions should be as close as possible to reality. So the technique and method of forecasting are necessary [6].

The two types of forecasting methods based on historical data are as follows: 1) Forecasting Technique, this forecasting method uses the experience and judgment of the forecasters. Those who have good knowledge of the subject can guess and 2) Official Forecasting Technique, forecasting this method relies on data to support and apply statistical knowledge. Mathematics to assist in data analysis. The prophet is everyone who understands. What to do and what steps to take Technical Analysis [7].

A. Complications in Diabetic Patients
Eye complications in diabetic patients (Diabetic retinopathy) depend on the duration of diabetes, blood glucose control (HbA1c), blood pressure control. And death rates caused by degenerative retina Diabetic patients with cataracts found 46.8 percent. Diabetic retinopathy is a major complication problem Partly due to the increasing prevalence of diabetes Primary retinal disorders which can maintain and prevent vision loss.
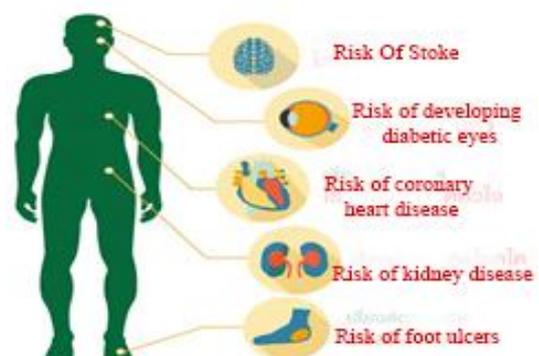

Figure 1. Thai Health (2018)

B. Kidney Complications from Diabetes (Diabetic Nephropathy)

Kidney disease in diabetic patients will increase dramatically in the future. From the down patients with renal replacement therapy by the Kidney Association Thailand in 2003 found that 38.8% were caused by diabetes and the patient rate. The higher the prevalence of renal failure, the five - year survival rate was 44 %, compared to 76% in non-diabetic subjects [cc]. Therefore, if there are no preventive measures, both screening and prompt treatment will cause loss of resources in the care of diabetic patients with renal complications is shown in Figure 2.
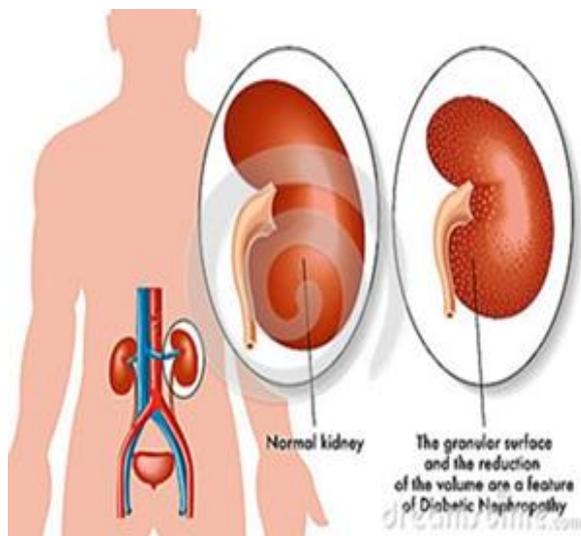

Figure 2. Manager (2013)

Neurological disorders are common in patients with complications of diabetes, neurological (Diabetic Neuropathy). The first phase may not be. The symptoms are so severe that it can affect the course of a patient's daily activities. And that is the risk factor. If you have had foot ulcers before, it can increase your risk by 35 times. Foot loss is one of the major risk factors for foot ulcers. And lead to cut leg. Foot examination is easy. There should be a campaign to provide more foot care for diabetics. In addition, due to geographic factors, and most Thai people still have agricultural occupations, Thai patients have ulcers are easily infected and spread.

Figure 3. NCD Office (2014)

In the past, traditional health analysis tools may not be complex or ease of use, which will provide more convenience. However, data nowadays from health service of the Ministry of Public Health locations nationwide are in forms of large data that need intelligence, analytic tools such as Hadoop / MapReduce, which is supporting the use of data analysis for Bigdata characteristics of 8Vs.

Figure 4. 8 Vs' Big Data Characteristic [8]

Data visualization is the translation of statistical data that people visually understand as a storyteller in which the audiences can easily understand [9]. The ability to work with

tools or programs is often used in ways that work across screens, widgets and screens. Help you design the dashboard as needed. Using a visualization tool as well and the selection of tools. Data visualization is Microsoft Power Bi for this research. Because it is a capable tool that is quite high Convenient and accepted.

The main objective of the image data analysis technique is to examine and understand the behavior of information in terms of storytellers [10]. In general, there are 4 analytical techniques in Big Data:

1) *Descriptive Analytics:* Organizations must analyze that data. To get a clearer and more detailed picture.

2) *Diagnostics Analysis:* Organizations must examine relational databases from multiple perspectives and sources. To identify relationships or patterns.

3) *Predictive Analytics:* analysis of current and past data to predict future or events that will happen which is more important.

4) *Prescriptive Analytics:* Collect all analyzes together to determine the best decision and possible practices.

III. PROPOSED MODEL

This research is semi-experimental research. Which aims to predict renal complications that occur in diabetic patients. The researchers used 6 million diabetic patient data from a large database. Which is the medical and health information warehouse of the Ministry of Public Health Through a qualitative process to identify, analyze and process (ETLs) according to Dataset to obtain data and interpret results according to relevant factors. The classification of data is determined by factors related to kidney complications in diabetic patients? In this article, the researcher uses Microsoft Power BI as a research tool to analyze, process and predict kidney complications in diabetic patients.

Based on Big Data analysis, The sample group contains information about diabetes

(DM) from 43 health data files nationwide with Big Data from the fiscal year 2013 to 2018, approximately 5 million recorded in experimental form. In data analysis and design, it may be necessary to prepare Because in Big data. The sample size used may not be complete, accurate and complete. Clear data to get quality information. However, the researcher will use this amount of information in large numbers. It is useful in predicting kidney complications of diabetes. To reduce complications that occur in diabetic patients which affects future death.
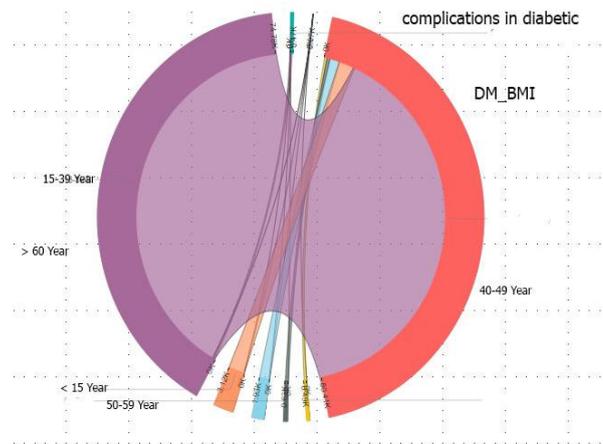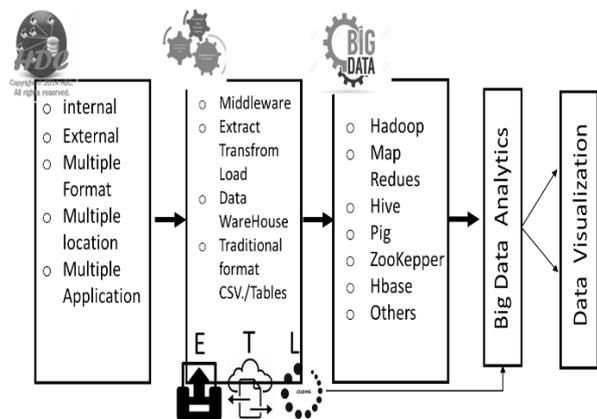

Figure 5. Data Visualization Model

Figure 6. An Applied Conceptual Architecture of Big Data Analytics.

The factors that are used as criteria for finding risk opportunities in this research are age (in years), gender-determined to 1 = male, 2 = female, body mass index (BMI) classified as 1) lower than 18.5 (thin 2) 18.5 - 22.9 (normal) 3) 23.0 - 24.9 (overweight) 4) 25.0 - 29.9 (obese level 1) 5) equal to or greater than 30 (obese, level 2) and smoking history

separated into 0. No smoking 1. Smoking, level HbA1c 1) lower than 7 2) greater than 7, if more than 7, it is a high-risk opportunity, Blood pressure Upper and lower body do not exceed the standard 140/90 mmHg, the Total Cholesterol Value <200 = Normal, Triglyceride Value <150 = Normal, LDL Cholesterol Value <130 = Normal, HDL Cholesterol Value <50 = Normal Not exceeding the standard, the level of protein in urine. The result is 0 = Negative 1 = Trace 2 = Positive, the duration of diabetes is (year) <1,1–2,3–5,6–10, > 10, Family history of diabetes, diabetic retinopathy Because of various factors These will have an impact on the chances of kidney complications in diabetic patients.

TABLE I
THE RESULTS OF FACTORS AFFECTING KIDNY COMPLICATION IN DIABETIC PATIENTS

| Factors Affecting Kidny Complication	Value	Factor
Sex	1,2	Male, Female
Age	Year	<15,15-39,40-49,50-59,60 Up
BMI	Value	<18.5=Underweight, 18.5 – 22.9=Normal, 23.0 – 24.9=Overweight, 25.0 – 29.9=Obese I, =>30=Obese II
Smoking	0,1	No smoke/smoke
HbA1c	Value	Normal<7, =>7 Risk
BP	Value	>140 SBP/>90 DBP=Risk
Total Cholesterol	Value	<200=Normal, >200 Risk
Triglyceride	Value	<150=Normal, >150 Risk
LDL Cholesterol	Value	<130=Normal, >130 Risk
HDL Cholesterol	Value	<50=Normal, >50 Risk
Micro/Macro Protein	Value	0=Normal, 1=Trace, 2=Positive
Diabetes duration	Year	<1,1–2,3–5,6–10,>10
Family history of diabetes	0,1	Normal/Risk
Heart disease	0,1	Normal/Risk
Diabetic Retinopathy	0,1	Normal/Risk

The development of a database structure for the analysis of renal complications in diabetic patients uses cube format Contains Fact tables and Dimension tables in the Star Schema format using attributes based on Data set as a Classification.

TABLE II
THE FACTORS AFFECTING KIDNY COMPLICATION IN DIABETIC PATIENTS

ID	AGE	Sex	HbA1c	BP	Creatinine	LDL	HDL	HD	Protein	BMI
1	45	2	10.5	160/95	2.03	0	0	1	0	1
2	65	2	8.3	110/70	1.25	0	0	1	0	2
3	61	2	6.3	140/100	0.94	144	0	1	0	2
4	42	2	7.5	120/90	1.20	0	1	1	1	4
5	77	1	6.0	135/99	0.56	0	0	2	1	2
6	77	2	6.5	150/118	1.19	200	1	2	1	3
7	43	1	7.9	165/90	0.7	194	0	1	1	1
8	42	1	9.5	130/80	0.56	0	0	1	1	2
9	40	2	10.2	155/95	0.63	0	1	1	0	5
10	73	2	8.0	130/90	0.85	0	0	1	0	4

In the process of making an analysis has used all diabetic patients from 2013-2018 from Big Data system Come to analyze the relationship and make Decision Trees according to variables to identify by chance of causing kidney complications in diabetic patients.
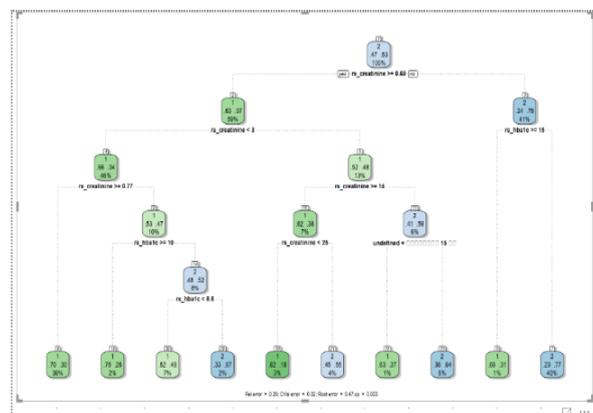
Figure 7. The Decision Trees System for Kidney Complications in Diabetic Patients on Semantic Model

IV. THE EVALUATION

Based on this data, there were 6.9 million records of diabetic patients in Thailand. Most diabetic patients are female. Accounted for 70 percent of the average age of 59 years (standard deviation 10.7) and 26 percent of kidney disease patients with diabetes when

controlling other variables. Factors affecting complications of diabetic nephropathy at a significant level 0.05 Sex, age, body, body mass index (BMI), HbA1c levels, triglyceride levels, LDL levels, HDL levels, total cholesterol levels, urinary protein levels, serum creatinine levels, family history of diabetes, Hotels smoking, duration of diabetes. Diabetes is a common condition in patients with diabetes. If the patient is able to control HbA1c levels, blood lipids, urine protein levels, and the EGFR assessment results are good. Will help slow down the deterioration of the kidneys and kidney failure.

In the assessment, the Bootstrap Decision tree is used to accurately predict renal complications in diabetic patients. Using data from Big Data system with diabetic patients who have passed the classification of information from patients in the whole country which, come to receive services in approximately 6 million records and then classified by risk factors, according to relevant variables To identify and assess the likelihood of kidney complications According to the target group, time and place to compare with past chronic kidney disease patients.

TABLE III
THE RESULTS OF MODEL PERFORMANC

No.	Categories	Precision	Recall
1	*Diabetic Patient*	95.5%	89.5%
2	*Kidney Complication*	94.7%	83.7%
3	*Factors Affecting*	91.2%	87.2%
Overall Performance		**93.8%**	**86.8%**

V. CONCLUSIONS

In this article, we propose a model for the use of data visualization tools. To predict renal complications in diabetic patients From a large database system By passing through the data mining process ETLs and classification based on data sets that are datasets According to the objectives of the research as follows: Many variables affecting kidney complications in diabetic patients And can tell that How many people with diabetes have the chance of having kidney complications that much likely higher risk of HbA1c or less than 7 and has positive or negative effects of protein in the urine as well as laboratory results not exceeding the standard There will be no chance or chance of having kidney complications. And when examining the historical data back to the group with chronic kidney disease, it was found that the factors that were found and matched with this research.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Key message. [2011, Nov 16] Retrieved from http://www.idf.org/worlddiabetesday/2009-2013/key-messages.

[2] (2009). Public Health Statistics, 2009. Bureau of Policy and Strategy. Bangkok: Veterans Organization Publishing House.

[3] Public Health Statistics, 1998-2009. Bureau of Policy and Strategy Office of the Permanent Secretary for Public Health.

[4] Arbsuwan, N. & Phanwate, N. (2012). The World Diabetes Campaign fiscal year 2012. Bureau of Non-Communicable Diseases Department of Disease Control, Ministry of Health.

[5] Bureau of Policy and Strategy Office of the Permanent Secretary of the Ministry of Public Health. NCD report 1 fiscal year 2011 (according to the 1st NCD report form 1 - 2) in the Royal Water Reverence Project Caring for public health.

[6] Rensom, K. (2002). Sale forecasting. Bangkok: Chulalongkorn University.

[7] Ruen, K. (2002). Sale forecasting. Bangkok: Chulalongkorn University.

[8] Rajeshwari, D. (2015). State of the Art of Big Data Analytics: A Survey. International Journal of Computer Applications (0975-8887), 120(22).

[9] Heckerman, D., Chickering, D.M., Meek, C., Rounthwaite, R., & Kadie, C. (2000). Dependency networks for inference, collaborative filtering, and data visualization. Journal of Machine

Learning Research, 1, 49-75.

[10] Stewart, R. & Davis, K. (2016). Big Data. In Mental Health Research: Current Status and Emerging Possibilities, 51(III), 1055-1072.