International Conference on Computational Science and Information Management (ICoCSIM)
Vol. 1 (2012) 140-146

The long-term prediction of Type II Diabetes Mellitus: A Review Study

Amatul Zehra[a], Tuty Asmawaty[a]*, Sabira Khatun[b]

[a] *Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan,Pahang 26300, Malaysia*
[b]*School of Electrical System Engineering, Universiti Malaysia Perlis,Perlis Malaysia*

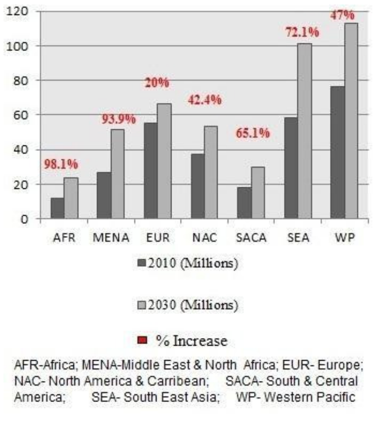
Abstract

Type II Diabetes Mellitus is one of the silent killer diseases worldwide. According to the World Health Organization, 346 million people are suffering from diabetes worldwide. In this paper, we reviewed some models and systems used for diabetes prediction, comparing and exploring various related research works. We found some models and systems to predict and diagnose type II diabetes mainly for a short term. This led to the related open issues of the development of a model for the prediction of type II diabetes in the long run. We analysed the need of a relation between the major factors that lead to the development of diabetes. We aim to study the pattern behaviour in existing data. Furthermore, we have proposed to classify the data using k-means algorithm and the prediction through particle filter method. The model would be able to do long term prediction for potential persons. This would be beneficial to overcome the sharp rise of diabetes globally.

*Keywords:*Diabetes prediction; Type II Diabetes Mellitus

1. Introduction

Diabetes Mellitus has become a common health problem nowadays, which would affect people and lead to various disablements like cardio vascular disease, visual impairments, leg amputation and renal failure if diagnosis is not done in the right time [1]. Diabetes can affect people due to the lack of insulin in the blood. Insulin is a natural hormone secreted by thepancreas, which acts as a key to unlock the body cells so that sugar, starch and food molecules can be absorbed and hence be utilized by the cells to generate energy required for daily life. Insulin deficiency is due to either of the two conditions. First is when the pancreas does not produce insulin at all. This leads to type I diabetes mellitus (T1DM) which is usually found by birth. Second state is when the body does not respond correctly to the insulin produced by the pancreas and hence the glucose that is consumed by the person is locked inside the blood instead of entering into the cells of the body. This ineffective insulin leads to type II diabetes mellitus (T2DM). Among these, type I diabetes is usually diagnosed in children and type II is the most common form which affects adults [2].



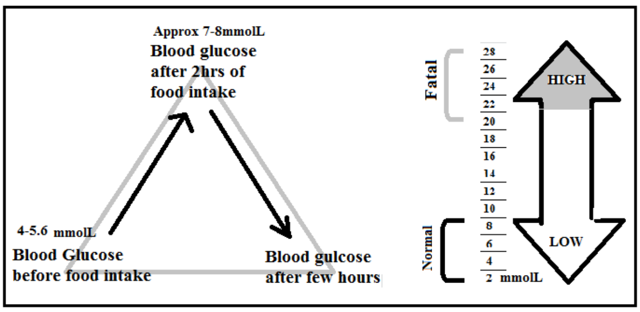


Fig. 1.(a) Region-wise estimated rise in diabetics by 2030 (Diabetes Atlas 4th Edition, International Diabetes Federation);(b) Ranges of normal and high blood glucose levels (www.diabetes.co.uk)

1.1. Diabetes-A Global Threat

The International Diabetes Federation has estimated an alarming rise in the number of diabetics by the year 2030, Figure 1(a) [3]. A sharp rise in diabetics has been observed in Asian region with 138 million Asians including 14.9% Malaysians [4]. From

1996 to 2006, the number of diabetics in Malaysia had increased by almost 80% and reached to 1.4 million adults above the age of 30. Among those, almost 36% were undiagnosed, resulting in complications that required more intensive medical care, putting great strain on the existing overstretched health services [5].

This paper focuses to investigate the possible solutions for the group of people who are at a risk of developing type II diabetes in future. We aimed to study type II diabetes because this type can be prevented by adopting proactive measures. Our team of researchers proposed to develop a prediction model based on existing data. For this purpose, we intend to use Pima Indian diabetes dataset. Eventually, the model would be able to answer about the need for significant and urgent requirement to: (i) stop sharp rise in diabetes, (ii) grow public health awareness, and (iii) prevent the onset of this disease.

The paper is organized as follows: section 2 gives a brief overview on type II diabetes followed by the review of the prediction and diagnostic models related to diabetes. Section 3 is the proposed study of this review paper and in the end is the conclusion.

2. Type II diabetes

Type II diabetes is sometimes called non-insulin dependent diabetes or adult-onset diabetes [3].At least 90% of all cases of diabetes are victims of this type. It strikes a person due to insulin resistance and relative insulin deficiency, either of which may be present at the time that diabetes becomes clinically evident. The diagnosis of type II diabetes usually occurs after the age of 40 but can occur earlier, especially in populations with high diabetes prevalence. Type II diabetes can remain undetected for many years and the diagnosis is often made from associated complications or incidentally through an abnormal blood or urine glucose test. It is often, but not always, associated with obesity, which itself can cause insulin resistance and lead to elevated blood glucose levels.

The normal range of fasting blood glucose level is between 4.0-5.6 mmol/L. After consuming a meal, the blood glucose level rises in the blood and can reach up to 7.8mmol/L. Any value higher than these ranges indicates the prevalence of diabetes. After two hours of having a meal, the blood glucose level drops again, Figure 1(b) [6]. There is also a condition called pre-diabetes. It is that state, where the blood glucose level is higher than the normal range but not high enough to be stated as diabetes.

Individuals can be categorized into three groups namely, ‘healthy, ‘pre-diabetics’ and ‘diabetics’. Blood glucose levels for all three categories vary accordingly. Table 1 shows the normal and post-meal ranges of blood glucose levels for these groups.

Table 1. Fasting and post meal normal ranges of blood glucose. (www.diabetes.co.uk)

Individuals	Fasting	2 hours after a meal
Healthy	4.0 - 5.6 mmol/L	<7.8 mmol/L
Pre-diabetics	5.6 - 7.0 mmol/L	<7.8 mmol/L
Diabetics	> 7 mmol/L	≥ 7.8 mmol/L

The need for avoidance and better management of type II diabetes has been an important issue since ages. Medical practitioners and researchers have investigated and continue to find solutions to overcome this disease. Various researches and studies are done on predicting the blood glucose levels for type II diabetes patients for a short term. Most of the predictions helped to decide the diet control and physical activities in order to maintain a healthy life [7].

3. Review of the type II diabetes prediction and diagnosis models

Due to rising cost of health care, it is useful to assist patients to control diabetes by themselves. In many instances, early information related to diabetes might help in avoidance, curing and appropriate treatment of the disease.Many computer programs or systems were developed and are being developed by emulating human intelligence that could be used to assist the users or patients in managing diabetes [8].We assessed different systems such as artificial intelligence systems, mobile phone applications and specially designed devices for the prediction and diagnosis of diabetes. The focus of this paper is to investigate for a model to predict and diagnose diabetes in the long run. Most of the models have been developed to diagnose diabetes and predict the blood sugar level for a short term. However, according to the authors’ knowledge, there are rarely any systems developed to predict the onset of diabetes in the long run.In the next section, a brief review on all related systems is done.

3.1. Models based on artificial intelligence

Artificial intelligence (AI) is a branch of computer science that aims to create machines that work logically. Literatures show that AI has been a great source of diabetes prediction and diagnostic tools. A short overview is shown in table 2.

Table 2. Review of artificial intelligence based models

	Technique used	Parameters used	Results obtained
1	• Artificial Neural Network (ANN) is used for the prediction • Back propagation algorithm and supervised training method are used. • 8-input and 4-input procedures are compared.	• Number of times pregnant • Plasma glucose concentration at 2 hours in an OGTT • Diastolic blood pressure • Triceps skin fold thickness • 2-hour serum insulin • BMI • Diabetes pedigree function • Age	• The comparison between the 8-input and 4-input procedures showed that the 8-input method gives higher performance. • 8-input procedure took 790 iterations for an accuracy goal of 10^{-6} while the 4-input procedure took 1010 iterations for an accuracy goal of 10^{-6}.

2	• Artificial Neural Network based classification model to predict whether a person is suffering from diabetes.
	• Genetic Algorithm is used for feature selection.
	• Comparison is done with other models such as Functional link ANN, Nearest Neighbour (NN), k-nearest neighbour (kNN), nearest neighbour with backward sequential selection (BSS) of feature, multiple feature subset (MFS1) and MFS2.
	• Pima Indian Diabetes dataset is used.
		• Number of times pregnant
		• Plasma glucose concentration at 2 hours in an OGTT
		• Diastolic blood pressure
		• Triceps skin fold thickness
		• 2-hour serum insulin
		• BMI
		• Diabetes pedigree function
		• Age
		• Class variable (0 or 1)
			• ANN model performs better than the models compared.
			• The following results were shown
			  • NN-65.1%
			  • kNN-69.7%
			  • BSS-67.7%
			  • MFS1-68.5%
			  • MFS2-70.5%
			  • Novel ANN- 73.4%
			  • FLANN- 59.8%

3	• An alternative method was used to overcome the missing values in a dataset. • By using the back propagation algorithm, missing value analysis techniques and pre-processing methods were analyzed.	• Pima Indian diabetes dataset was used. (Same as above)	By adopting the modifications, they achieved 99 % classification accuracy results.
4	• A supervised multilayer feed-forward network with back propagation learning algorithm was used. • Two different ANN architectures were developed. • One with single hidden layer and the other with double hidden layer.	• Random blood sugar test result • Fasting blood sugar test result • Post plasma blood sugar test result • Age • Sex • Occupation	• The results showed that the network is able to classifydiabetic and non-diabetic patients with the networkperformance of 92.5% • It also showed that by increasing the number of hidden layers as well as neurons for a given learning rate, least number of iterations was taken for attaining the error goal.
5	• A fuzzy estimator was developed for estimating hypoglycemic (dropping of blood sugar level) condition. • It was a non-invasive method by identifying different states of the patient	• Heart beat • Skin impedance	An excel macro has been written which takes heart beat rate and skin impedance as input and predicts the index of hypoglycemia.
6	• Feature extraction method was used to predict near future (30 minutes) blood glucose readings for diabetics. • Diabetic Dynamic Model was developed to extract knowledge based on the food intake. • ANN was used to predict the blood glucose levels.	• Blood glucose test results for 24 hours.	When compared with a different model AIDA, this model gave lesser root mean square error (RMSE).

A method for diagnosing diabetes was proposed by Jaafar and Ali using back propagation neural network algorithm as shown in Table 2[9]. The inputs to the system were plasma glucose concentration, blood pressure, triceps skin fold, serum insulin, Body Mass Index (BMI), diabetes pedigree function, number of times a female person was pregnant and age. The biggest challenge was the missing values in the data set in this study. In another study, an artificial neural network based classification model was developed to predict whether a person is suffering from diabetes [10]. Genetic algorithm is used for feature selection. Comparison is done with other models while using the open source database, Pima Indian Diabetes dataset. The comparison showed that ANN model performs better than the models compared.This system has been modified and presented by Jayalakshmi and Santhakumaran, to overcome the missing values that were not considered by the previous system [11]. They constructed the data sets with reconstructed missing values to increase the classification accuracy. They also proposed an alternate method to overcome missing value by performing data preprocessing. This sped up the training process by reducing the actual learning time. Various missing value techniques and pre-processing methods were analyzed by them to improve the results classification accuracy (99%).

Dey, et al. proposed a method to predict T2DM using back propagation algorithm of Artificial Neural Network (ANN) [12]. The problem of diagnosing diabetes was treated as a binary classification, i.e., those predicted to be diabetic fall under category 1 and others fall under category 0. The basic architecture of ANN used for accomplishing this classification task was a supervised multilayer feed-forward network with back propagation learning algorithm. The considered parameters in this system were the results of random blood sugar test, fasting blood sugar test, post plasma blood sugar, age, sex and their occupation. The performance was in terms of absolute error calculation between network response and desired target.

A fuzzy estimator was developed for estimating hypoglycemic condition [13].Hypoglycemia is a condition when a diabetic patient's blood glucose level drops very low. It was a non-invasive method by identifying different states of the patient like heart beat and skin impedance. Based on these two inputs an excel macro has been written. The heart beat rate and skin impedance is then put in as input to the system and it predicts the index of hypoglycemia.

Feature extraction method was used to predict near future (30 minutes) blood glucose readings for diabetics [14]. Diabetic Dynamic Model was developed to extract knowledge based on the type of the food intake.ANN was used to predict the levels. The model used the blood glucose test results as the input parameters.When compared with a different model AIDA, this model gave lesser root mean square error (RMSE).

3.2. Models based on smartphone applications

Chemlal et al. in 2011 developed a real time predictive system which was embedded in an application HealthiManage [15]. The application was used on an iPhone to predict a short term blood glucose reading. The system consisted of a prediction algorithm and several machine learning and curve fitting methods. It used exponential distribution functions with estimation. It had a built-in accelerometer on the iPhone which is used for activity recognition (type, duration and intensity). The application provided relevant feedback to the patient after every glucose input reading comparing the measured and predicted readings. Moreover, the application monitored physical activity and adjusts the predictions accordingly. It aided the patients to control their glucose levels by suggesting them to cut down on sugar and carbohydrates and increase their activity.

3.3. Models based on other techniques

Apart from the above mentioned techniques, there are a few more systems developed which use different methods and procedures. They are reviewed in the following table.

Table 3. Models based on other techniques

	Technique used	Parameters used	Results obtained
1	• A system to predict short term blood glucose level using equal time interval, local fuzzy reconstruction method and minimal linear method. • Two types of insulin therapies are used and compared. • Based on the prediction, an appropriate amount of insulin shot is estimated.	• Blood Glucose Concentration • HBA1C index value • Amount of insulin administered at bedtime	• The results indicated that based on the predictions, insulin therapy 1 and insulin therapy 2 was used. • In insulin therapy 1 the change of frequency was 52.2% and the correctness of glycemic control was 84.5%. • In Insulin therapy 2, the change of frequency was 32.9% and the correctness of glycemic control was 99.2%.

2	• An algorithm was created that predicts blood glucose levels after a meal intake. • This prediction would help the patient to decide the amount of insulin to be taken.	• Post meal blood glucose level. • Plasma glucose distribution volume per unit of body weight.	• Based on the algorithm convincing prediction of blood glucose level was obtained. • Necessary insulin shots were administered. • The blood glucose level was then decreased to an acceptable range.
3	• Data mining by RapidMiner for diabetes data analysis and diabetes prediction model	• Using the Pima Indian Diabetes dataset • Number of times pregnant • Plasma glucose concentration at 2 hours in an OGTT • Diastolic blood pressure • Triceps skin fold thickness • 2-hour serum insulin • BMI • Diabetes pedigree function • Age • Class variable (0 or 1)	• A decision tree was used for prediction with 72 % of accuracy. • ID3 Algorithm was also used for this purpose which gave 80 % accurate results.
4	• A Universal data driven model is proposed based on glucose data from a diabetic applied to predict the glucose concentration of other patients. • Three different Continuous Glucose Monitoring devices were used. • Root-mean-squared error and Clarke error grid analysis methods were used for comparisons.	• Age • Type of diabetes • Insulin user or not • BMI • HBA1C value • Pregnant or Lactating	The result was 99 % accurate and the sensor-predicted glucose concentrations lay in the acceptable zone.
5	• Based on the past glucose data, with time-varying parameters, 30 min ahead glucose level can be predicted. • A Continuous Glucose Monitoring (CGM) device is used.	• Readings from CGM devices were used. • Two methods were applied • First-order polynomial • First-order autoregressive model	Results indicated that glucose can be predicted ahead in time and the patient can be administered insulin or sugar respective to the predictions.

As shown in Table 3, there are several different types of systems developed over the years apart from artificial neural network. A system was developed to predict short term blood glucose level using equal time interval, local fuzzy reconstruction method and minimal linear method [16]. Two types of insulin therapies are used and compared. Based on the prediction, an appropriate amount of insulin shot is estimated.The results indicated that based on the predictions, insulin therapy 1 and insulin therapy 2 was used. In insulin therapy 1 the change of frequency was 52.2% and the correctness of glycemic control was 84.5%. In Insulin therapy 2, the change of frequency was 32.9% and the correctness of glycemic control was 99.2%.

An algorithm was created that predicted blood glucose levels after a meal intake [17]. This prediction would help the patient to decide the amount of insulin to be taken. The parameters used were post meal blood glucose level, plasma glucose distribution volume per unit of body weight. Based on the algorithm, convincing prediction of blood glucose level was obtained and necessary insulin shots were administered. The blood glucose level was then decreased to an acceptable range.

Han et al. used data mining techniquethrough*RapidMiner* for diabetes data analysis and diabetes prediction model [18]. A decision tree was used for prediction with 72 % of accuracy. ID3 Algorithm was also used for this purpose which gave 80 % accurate results. They used the open source database Pima Indian Diabetes dataset and the inputs were the same as were used in [10]

A universal data driven model is proposed based on glucose data from a diabetic applied to predict the glucose concentration of other patients [19]. Three different Continuous Glucose Monitoring devices were used. Root-mean-squared error and Clarke error grid analysis methods were used for comparisons.
The parameters used were age, type of diabetes, insulin user or not, pregnant or lactating, BMI and HBA1C value. The result was 99 % accurate and the sensor-predicted glucose concentrations lay in the acceptable zone.

In another system, 30 minute ahead glucose level can be predicted based on the past glucose data, with time-varying parameters [20]. A Continuous Glucose Monitoring (CGM) device is used for this work. Readings from CGM devices were used. The researchers applied two methods

- First-order polynomial
- First-order autoregressive model

Results indicated that glucose can be predicted ahead in time and the patient can be administered insulin or sugar respective to the predictions.

4. Proposed prediction model based on cognitive pattern recognition and particle filter

Many researches are done on the prediction and diagnosis of diabetes for the short term [21][22]. After the literature survey, we have found the need to develop a model that predicts the onset of diabetes in the long run. By long term prediction we mean that the prediction is done well before the onset of the disease so that the potential patient is able to carry preventive measures to overcome or escape the disease.

This model is proposed by considering those people who are at a risk of developing diabetes in future. We have investigated to develop the model by using cognitive pattern recognition on the Pima Indian diabetes dataset which is also used in [9][10][11][18]. Pattern recognitionis often used to optimally extract patterns based on certain conditions and to separate one class from the others [23]. K-means classification algorithm would be used to classify the data. Furthermore the prediction would be done by using particle filter method on the developed model. Hence the objectives of this proposed study are

- To identifythe patterns in the dataset.
- To classify the data into respective classes or groups.
- To do the long term prediction of diabetes.

5. Conclusion

In this paper, various investigations on prediction and diagnosis of type II diabetes mellitus are presented. There are many prediction systems for type II diabetes mellitus but they are mainly for short-term predictions. Therefore the open issue is to build a long-term prediction model which would be beneficial for the persons who are at a high risk of developing diabetes and are unaware of this fact. This type of system can produce more accurate results and will be helpful to reduce and prevent diabetes growth globally.

References

1. http://www.who.int/topics/diabetes_mellitus/en/, (Last access date: 30th September 2012
2. Pobi S, Hall LO. Predicting juvenile diabetes from clinical test results. Proceedings of the 2006 International Joint Conference on Neural Networks (IJCNN) 2006. 2159 – 65.
3. http://www.idf.org/diabetesatlas/5e/regional-overviews, (Last access date: 30th September 2012)
4. 'Sharp rise of diabetics in Asia', The Malay Mail, 3rd December 2010, (Last access date: 13th October 2011).
5. 'Alarming rise in number of diabetics in Malaysia', The Star, 11th January 2010, (Last access date: 12th October 2011).
6. Normal and diabetic blood sugar level ranges. URL: www.diabetes.co.uk/diabetes_care/blood-sugar-level-ranges.html, (Last access date: 14th October

2011

7. Lauritzen JN, Arsand E, Vuurden KV, Bellika JG, Hejlesen OK, Hartvigsen G (2011). Towards a mobile solution for predicting illness in type 1 diabetes mellitus: Development of a prediction model for detecting risk of illness in type 1 diabetes prior to symptom onset. Proceedings of the 2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronics Systems Technology (Wireless VITAE), 1-5
8. Barakat NH, Bradley AP, Barakat MNH. Intelligible support vector machines for diagnosis of diabetes mellitus. Proceedings of the IEEE transaction on information technology in Biomedicine 2010; 14:4.
9. Jaafar SF, Ali DM. Diabetes mellitus forecast using artificial neural networks. Proceedings of the Asian conf. of paramedical research 2005
10. Pradhan M, Sahu RK. Predict the onset of diabetes disease using Artificial Neural Network. *Intl J Comp Sci & Emerging Tech* 2011;2:303-11
11. Jayalakshmi T, Santhakumaran A. A novel classification method for classification of diabetes mellitus using artificial neural networks. Proceedings of the 2010 International Conference on Data Storage and Data Engineering 2010
12. Dey R, Bajpai V, Gandhi G, Dey B. Application of artificial neural network technique for diagnosing diabetes mellitus. Proceedings of the 2008 IEEE Region 10 Colloquium and the Third ICIIS 2008
13. Hastings G , Ghevondian N, Nguyen H. A self-organising fuzzy estimator for hypoglycemia monitoring in diabetic patients.Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1998;20 No.3
14. Eskaf K, Badawi O, Ritchings T. Predicting blood glucose levels in diabetics using feature extraction andArtificial Neural Networks
15. Chemlal S, Colberg S, Satin-Smith M, Gyuricsko E, Hubbard T, Scerbo MW, McKenzie FD. Blood glucose individualized prediction for Type 2 diabetes using iPhone application. Proceedings of the 2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC) 2011:1-2.
16. Katayama T, Sato T, Minato K. A blood glucose prediction system by chaos approach. Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEMBS 2004;1:750-3
17. Islam MS, Leech J, Lin CCY, Chrostowski L. Peak blood glucose prediction algorithm following a meal intake. 2007;579-82
18. Han J, Rodriguze JC, Beheshti M. Diabetes data analysis and prediction model discovery using RapidMiner. 2008 Second International Conference on Future Generation Communication and Networking 2008:96-9
19. Gani A, Gribok AV, Lu Y, Ward WK, Vigersky RA, Reifman J. Universal glucose models for predicting subcutaneous glucose concentration in Humans. Proceedings of the IEEE Transactions on Information Technology in Biomedicine 2010; 14: 157-65
20. Sparacino G, Zanderigo F, Corazza S, Maran A, Facchinetti A, Cobelli C. Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series. IEEE transactions on biomedical engineering 2007;54:931-37
21. Cescon M, Johansson R . Glycemic trend prediction using empirical model identification. Proceedings of the Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference 2009; 3501-6.
22. Ionescu C, Zlate S, De Keyser R. An approach to control the blood glucose level in diabetic patients. Proceedings of the 2006 IEEE International Conference on Automation, Quality and Testing, Robotics 2006;1: 48-53.
23. Zheng L, He X. Classification techniques in pattern recognition. Proceedings of The 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision WSCG2005.