AMODEL TO PREDICT AND PRE-TREAT DIABETES MELLITUS

Arockiam L[1], Dalvin Vinoth Kumar[2]
1,2Department of Computer Science, St. Joseph's College (Autonomous), Thiruchirapalli, Tamil Nadu, India

Abstract:

Diabetes Mellitus (DM) is one of the non-communicable diseases and it causes major health problems. According to a study, there will be 552 million diabetic patients by 2030 all over the world. The Things embedded with sensors that are connected to the internet is referred as Internet of Things (IoT). The collection, storage and analysis of data from IoT devices facilitate effective monitoring diabetic patients. In this paper, a model for prediction of diabetes is proposed. This prediction model consists of layer of sensors for data collection, layer for storage and layer for analytics. The diabetic data collection may include the data from other sources such as clinical experiments and questionnaire. The collected data are cleaned using pre-processing techniques. In the storage layer, the preprocessed data are stored in the warehouses. The predictive analytics is performed using statistical, data mining and machine learning algorithms in the analytical layer. This model provides an approach to predict the diabetic mellitus

Keywords—**Diabetes Mellitus, Predictive model, Diabetes predication, Clinical data analytics, Diabetics in India.**

I. INTRODUCTION

Diabetes is referred as diabetic mellitus in which blood sugar levels are too high. It is defined as a clinical syndrome characterized by hyperglycaemia, due to inadequacy of insulin in the human body. High levels of blood glucose can damage the blood vessels in kidney, heart, eyes and entire nervous system. Lack of awareness about diabetes can lead to these complications. According to WHO, India is the residence of the highest number of diabetics with the population of 79.4 million by 2030 [1]. The International Diabetes Federation said that nearly 52 % of Indians not aware that they are suffering from high blood sugar [2]. In particular, Madras Diabetes Research Foundation suggested that about 42 lakh individuals have diabetes and 30 lakh people are in pre-diabetes.

II. TYPES OF DIABETES MELLITUS

There are three types of diabetes. They are Type-1 diabetes, Type-2 diabetes and Gestational diabetes [3]. The presence of diabetics is identified using the following factors long-term blood sugar (HbA1C), fasting blood sugar, fotal triglycerides, Family history of high blood sugar, Waist measurement, Height and Waist-to-hip ratio[4].

i. **Type-1 Diabetes:** This type of diabetes is also called as insulin dependent diabetes. It will start from childhood. It is immune mediated and idiopathic forms of b cell dysfunction, which lead to absolute insulin deficiency [5]. This is also an auto-immune mediated disease process which gives rise to absolute deficiency of insulin and therefore total dependency upon insulin for survival. It increases the risk of heart disease and stroke. The symptoms are very thirsty, urinating frequently, rapid weight loss, feeling very hungry, feeling extreme weakness and fatigue, Nausea, vomiting and irritability. The treatments of type -1 diabetes are injections of insulin, oral medications or dietary modifications, physically activity, regular

check-up of blood sugar levels, controlling blood pressure and monitoring cholesterol levels [6].

ii. **Type-2 Diabetes:** It is also called as non-insulin dependent diabetes and adult onset diabetes. It is the most common form of diabetes. People may be affected by type-2diabetes at any stage, even during childhood [7]. Being overweight and inactive increases thechances of developing type-2 diabetes. It may originate from insulin resistance andrelative insulin deficiency. It can be controlled with weightmanagement, nutrition and exercise. The symptoms are very thirsty, urinating frequently,rapid weight loss, feeling very hungry, feeling extreme weakness, fatigue, nausea,vomiting, irritability, blurred vision, excessive itching, skin infections, sores that healslowly and dry and itchy skin [8]. Treatments such as using diabetes medicines, insulininjections, healthy food choices, exercise, Self Monitoring of Blood Glucose (SMBG),controlling blood pressure and monitoring cholesterol levels are some measures to control Diabetes Mellitus [9].

iii. **Gestational Diabetes:** According to the National Institutes of Health, the reported rate of gestational diabetes is between 2% to 10% of pregnancies [10]. Gestational diabetes usually resolves itself after pregnancy. It is caused by the hormones of pregnancy or a shortage of insulin. It causes risks to the life of baby which include abnormal weight gain before birth, breathing problems at birth, and higher obesity and diabetes risk later in life.

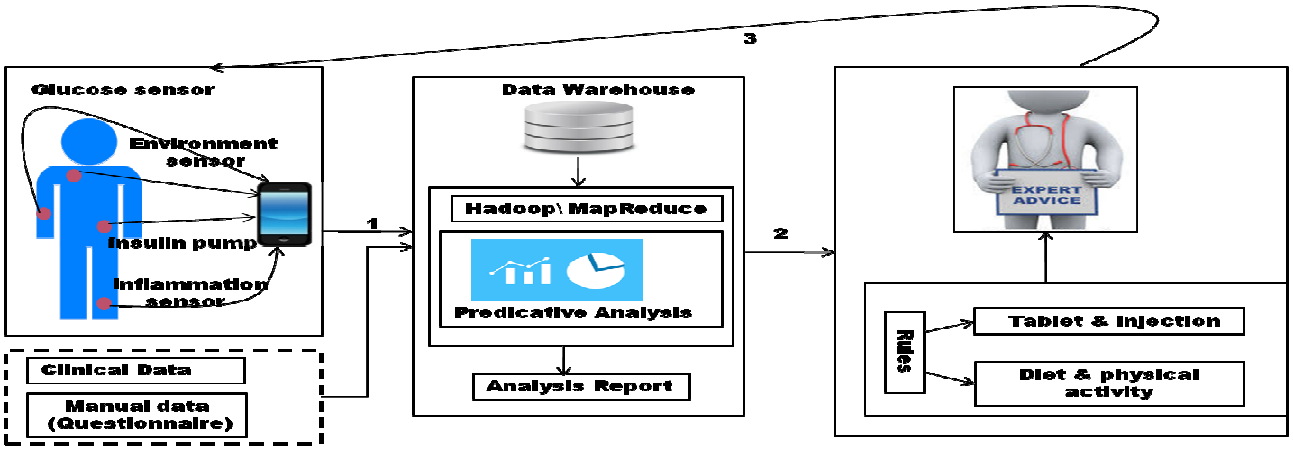
III. Problem Definition

In this modern era, human beings encounter different health issues. Most of the health issues are due to the food habits of the individuals. Based on the questionnaire, clinical data and sensor data, a predictive model is proposed to prevent the Diabetic Mellitus.

IV. The Proposed IDPM Model

A Diabetes Based Prediction Model plays animportant role in predicting diabetes and pre treatingdiabetic patients. It consists of three layers namelystorage layer and analytics and action layer. The layers in the proposed model for diabetic prediction are presented in Fig. 1.

The device layer consists of body sensors or wearables such as inflammation sensor, insulin pump, glucose sensor and environment sensor. These sensors are implanted in the body. These sensors are monitoring the health status of the human body.



Each sensor has its own capability to sense the data. The data gathered from these body sensors are electronic health record vitals, lab results and medical history of patients from hospitals or clinics. The second layer is cloud storage where the electronic records are stored securely. Cloud computing can be used for: data processing, data analysis and predictive analysis using statistical and data mining techniques and tools. Distributed Data analysis is performed by Hadoop/MapReduce.The predicative analytic engine sends report to doctors for consultation. The decision making at the cloud level is based on some rules such as diet pattern, physical fitness, current medicine intake etc. This layer is called as analytics and action layer. The rules for taking actions are set by physicians and medical experts for various health related issues. The rule based consultation will also consider previous health records and medical actions already taken. Here, the text pre-processing techniques are used. The doctor checks the analysis report of the patient. The doctor sends the treatment details from the prediction report to the insulin pump actuator as shown in Fig. 2.

Fig. 1. The proposed IDPM Model

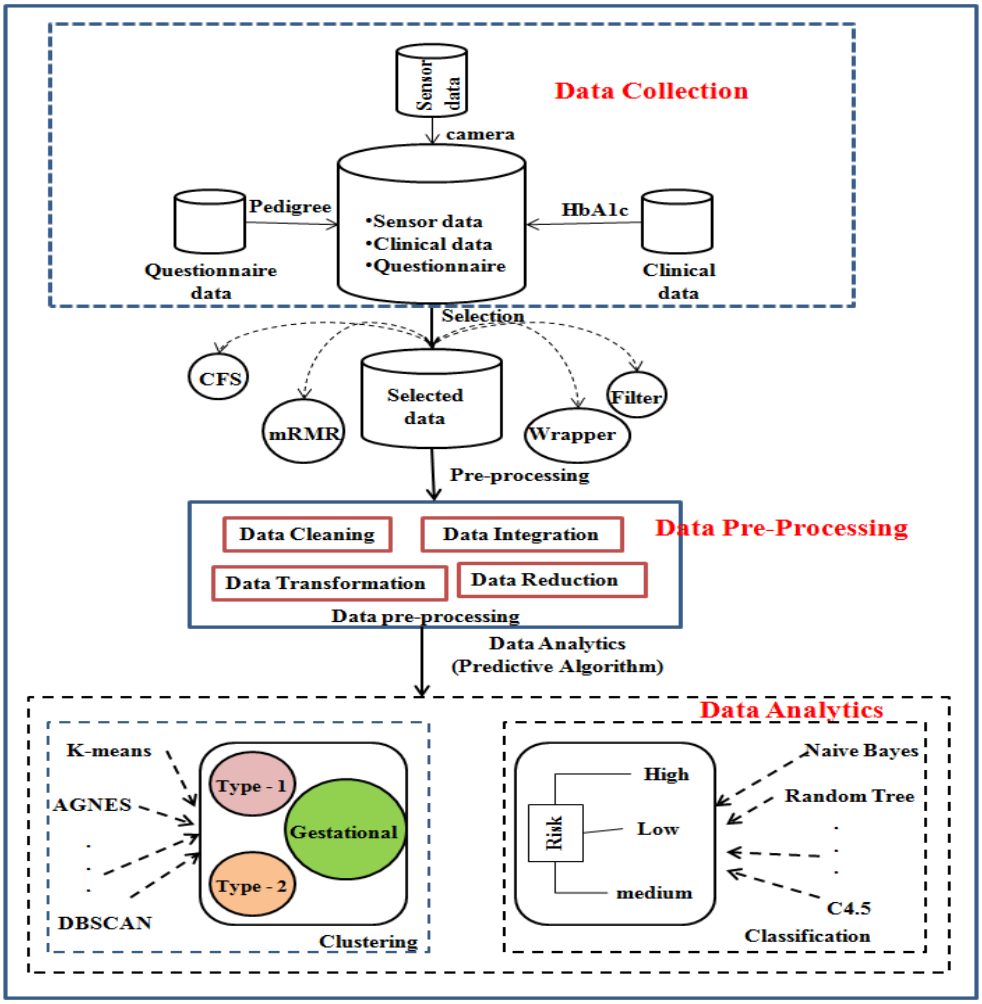


Fig. 2. The Diabetes Prediction

A. **Data collection**

Data collection is one of the most important stages of a research. Data collection is very demanding job which needs thorough planning, hard work, endurance, resolve and more to be able to complete the task successfully. Data Collection has two critical components. They are information gathering and decision making. Data collection is divided into two types. They are qualitative and quantitative data. Qualitative data are mostly non-numerical in nature. This means data will collect in the form of sentences or words. Quantitative data is numerical in nature and can be mathematically calculated.. In this proposed model, three types of data involved namely sensor data, clinical data and questionnaire data. The blood glucose level, body temperature, sleep time etc are collected from sensors. The clinical data like HbA1C test data are collected from clinical data. The family blood sugar history, number of time got pregnant etc., are collected from Questionnaire. Some of the data collected from sensors are qualitative and some are quantitative in nature.

B. **PRE-PROCESSING**

In real-time, it is very tedious to process massive amount of medical datasets containing information of individual patient health records so as to identify the disease pattern and to find out the causal association between them for planning curative actions. As a result, pre-processing of large volume of clinical data becomes essential. Data pre-processing is an important step in the data mining process. Data collection methods are largely loosely controlled, resulting in out-of-range values, missing values, etc. Observing the data which has not been carefully examined for such issues can produce misleading outcomes. If there is ample amount of incorrect and redundant information or noisy and unreliable data, then knowledge discovery becomes challenging.

Data pre-processing includes various steps such as cleaning, normalization, transformation, feature extraction and selection, etc. The outcome of data pre-processing is the complete data set with reduced attributes. The major drawback with clinical data set is the existence of redundant records. These redundant records cause the learning algorithm to be biased. So, eliminating redundant records is essential to enhance the detection accuracy. During pre-processing, the raw data is supplied as input and several suitable data pre-processing methods are applied thereby decreasing the invalid instances in the dataset.Data transformation and data validation are two important pre-processing techniques. They are explained below.

i. Data Transformation

In data transformation, the data are transformed or consolidated into forms appropriate for mining. Data transformation can involve Normalization, Smoothing, Aggregation and Generalization of the data.

a. *Normalization:* It is the process of converting the given values into a smallest range such as -1.0 to 1.0 or 0.0 to 1.0
b. *Smoothing*: Smoothing refers to removal of the noise from data. Smoothing techniques include binning, regression and clustering.
c. *Aggregation*: The Process of gathering information and expressing a summary form for the purpose of statistical analysis.
d. *Generalization*: Generalization is the process where low level or primitive data are replaced by higher level concepts through the use of concept hierarchies.

ii. Data validation

Data validation is defined as the assessment of all the collected data for entirety and reasonableness, and the elimination of error values. This step changes the raw data into validated data.. Data validation may be simple or complex depending on the way it is performed. Data validation can be updated either automatically or manually. The data validation helps to control the invalid data being entered into the system.

C. Predictive Analytics

The predictive analysis is carried out using classification and clustering algorithms. The clustering algorithms like K-means, DBSCAN etc. are used to cluster the pre processed data. The population is clustered based on urban/ village, male / female, educated / un-educated, diabetic / non diabetic etc. The clustered population is classified into DM high, low and medium using the classification algorithms like Bayesian network, J48, random tree etc. Some of the classification algorithms are explained below.

Bayesian network is statistical model which represents a set of variables and conditions. The relationship between the variables is carried out using Directed Acyclic Graph (DAG). In Bayesian network the nodes in DAG represent variables and edges between the nodes represent conditional dependency. Diabetic prediction using Bayesian network is influenced by the parameters. The Bayesian network for diabetic prediction with three variables (Blurred Vision (BV), Hunger and Fatigue (HF) and Family History (FH) of high blood sugar) implies Diabetes Mellitus (DM) with conditional dependency as shown in the fig. 3. The conditional dependency between the variables has two possibilities true and false respectively. The probability (pr) of occurrence of diabetics with respect to FS, BV and FV is given in equation 1.

$$Pr\,(FS\ BV\ FH) = Pr\left[\frac{FH}{(BVHF)}\right] * Pr\left[\frac{BV}{(HF)}\right] * Pr\,[HF]\ (1)$$

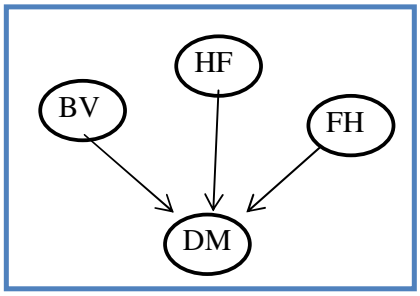


Fig. 3. Bayesian Network for Diabetes predication

The logistic regression is a statistical model used in various fields including machine learning. The decision is made by the influence of dependent variable (DV). In diabetic prediction, logistic regression utilizes general characteristics like age, Walking steps per day, Body Mass Index (BMI), Blood sugar level etc. Table 1 gives information about a set of patients (p1, p2, p3, and p4) walking steps per day and Blood sugar status. Logistic regression is suitable for the data in the table 1. The reason is, dependent variable Blood sugar status value is 1 or 0 represented to reduced or not reduced.The relationship between the steps walked per day and blood sugar status as shown in Fig. 4.

Table 1 Diabetes predication variable

Patient	Walking steps/day	Blood sugar Status
P1	4400	1
P2	3300	1
P3	2200	0
P4	0000	0

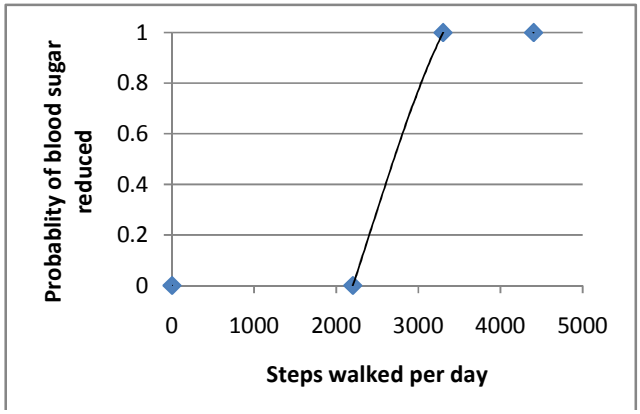


Fig. 4 Probability of blood sugar reduced versus number of steps walked.

V. Conclusion

Diabetes Mellitus is a chronic non communicable disease which has impact on human life span. A lot of datais collected from diabetic patients using IoT devices, clinical experiments and questionnaire, etc. The doctors can find value and make decisions when analysing the data accumulated from these sources. Early prediction of the deficiency will help the doctors to decide the treatment methods. Hence, a model for prediction of diabetes is proposed enabling pre-treatmentof the patients. .

Acknowledgment

This research work is financially supported by University Grants Commission, Government of India, under the Minor Research Project scheme. Ref. No.: F MRF-6517/16 (SER)/UGC).

REFERENCES

1. NDTV Food Desk, Updated: November 14, 2017 13:06 IST, available at: https://www.ndtv.com/food/world-diabetes-day-2017-number-of-diabetics-to-double-in-india-by-2023-1775180

2. Olson, Brooke. "Applying medical anthropology:Developing diabetes education and prevention programs in American Indian cultures", American Indian Culture and Research Journal, Vol:23, No:3, 1999, pp: 185-203.

3. World Health Organization. "Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus.", 1999, pp. 17-21.

4. American Diabetes Association. "Diagnosis and classification of diabetes mellitus", Vol:37, No:1 , 2014, pp : 81-90.

5. Gavin III, James R, "Report of the expert committee on the diagnosis and classification of diabetes mellitus." Diabetes care, Vol: 20, No:7, 1997, pp :963-967.

6. Diabetes Control and Complications Trial. "Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes." The New England journal of medicine, Vol: 353, No: 25, 2005, pp: 26-43.

7. Alberti, G., Zimmet, P., Shaw, J., Bloomgarden, Z., Kaufman, F. Silink, M. Type 2 diabetes in the young: the evolving epidemic. Diabetes care,Vol:27,No:7,2004, pp: 1798-1811.

8. Recognising Type-2 Diabetes', Feb 2016 Available : https://www.healthline.com/ health/type-2 -diabetes /recognizing-symptoms, [Accesed : 15-jan-2018].

9. Czupryniak, Leszek, "Self-monitoring of blood glucose in diabetes: from evidence to clinical reality in Central and Eastern Europe—recommendations from the international Central-Eastern European expert group.",Diabetes technology & therapeutics , Vol:16, No:.7, 2014, pp: 460-475.

10. Alatab, S., Fakhrzadeh, H., Sharifi, F., Mirarefin, M., Badamchizadeh, Z., Ghaderpanahi, M. Larijani, B.,." Correlation of serum homocysteine and previous history of gestational diabetes mellitus.",Journal of Diabetes and Metabolic Disorders, Vol : 12, No :1,2004, pp: 34-36.

11. Gillman, M. W., Rifas-Shiman, S., Berkey, C. S., Field, A. E., &Colditz, G. A. "Maternal gestational diabetes, birth weight, and adolescent obesity".Pediatrics, Vol:111, No: 3, 2003, pp:221-226.