An IoT Based diabetic patient Monitoring System Using Machine Learning and Node MCU

Amine Rghioui, Assia Naja, Jaime Lloret Mauri, Abedlmajid Oumnad

1 Research Team in Smart Communications-ERSC Research Centre E3S, EMI, Mohamed V University in Rabat, Morocco
2 International University of Rabat, FIL, TICLab, Morocco
3 Integrated Management Coastal Research Institute, Universitat Politecnica de Valencia, 46370 Valencia, Spain.

rghioui.amine@gmail.com, naja.assia@gmail.com, jlloret@dcom.upv.es, aoumnad@emi.ac.ma

Abstract. Diabetic patient monitoring is a systematic method that provides us with detailed information about the diabetic patient. Diabetic patient monitoring systems play a significant role in monitoring the patient’s health, especially with the use of Internet of Things (IoT) devices. Diabetic patient monitoring systems are able basically to monitor diabetic patients and save some data about blood glucose level, body temperature, and location. The role of this system is not limited to patients monitoring, it can also classify data using machine learning techniques. Predictive analytic for diabetic patients is very important due to its ability to help diabetic patients, their families, doctors, and medical researchers to make decisions on diabetic patient treatment based on big data. This paper describes a new system for monitoring diabetic patients and discusses predictive analytics using four different machine-learning algorithms. The performance and accuracy of the applied algorithms are discussed and compared to choose the best one in terms of several parameters.

1. Introduction

The technologies of the Internet of objects and their applications have made a great development by becoming in these days more accessible and more available, allowing a large number of objects to be interconnected via the Internet in several fields that are the field of health, home automation, industrial manufacturing, etc [1]. In the field of intelligent health, there are several applications that aim to improve care and improve the quality of life of patients with chronic diseases. Using IoT, the mobile health service becomes more important as it plays a very important role in monitoring and controlling patients who suffer from chronic diseases such as cardiovascular disease and diabetes [2][3]. In the field of intelligent health and more precisely in the part of patient monitoring, we find that patient data is very useful. Indeed to realize an IoT application in this field, the one must have assured the recording of a
large amount of data collected by using measurements of the medical signs on the patients. Analyzes can be applied to identify people who need "proactive care" to avoid worsening their condition. For example, patients in the early stages of certain diseases (for example, heart failure often caused by certain risk factors such as hypertension or diabetes) should be able to benefit from preventive care thanks to big data [4]. To improve the health of patients, some of them shared their private details to save lives in return [5].

Nowadays, there are several chronic illnesses, for instance, heart disease, stroke, cancer, chronic respiratory disease, and diabetes. This is a dangerous disease that is lately becoming one of the leading causes of death in the world, and which requires a lot of careful monitoring to keep patients healthy. Diabetes are caused by insulin resistance, and insufficient insulin production can lead to either an increase or decrease in the level of glucose in the blood, therefore the main challenge of the diabetic patient is to maintain the glucose level stable within a specific interval. If they can no longer comply with these conditions, some patients require urgent care to avoid worsening [6].

The growth of diabetic patients in the world implied an increase the use of continuous glucose monitoring devices (CGM) to control diabetic patient, these devices become the new method of continuous monitoring. They provide real-time information about the glucose level.

In this article, we present an intelligent system for monitoring diabetic patients using Node MCU and Machine Learning algorithms. The MCU node is connected with the glucometer to periodically record the glucose level in a diabetic patient. Using this collected data, patients can be monitored remotely by caregivers (patients, researchers, and doctors). As a result, patients and doctors alike have to process multiple records and interpret the huge amount of data to adjust insulin doses and keep blood sugar as close to normal as possible. An intelligent algorithm is implemented in our system, which is able to send the data to caregivers, to store them in a database as positive or false positive after being validated (or not) by the doctor. In our case study, the included sensors facilitate the monitoring of diabetic diseases.

The rest of this paper is organized in the following: Section 2 presents the related work. Section 3 describes our proposed diabetic monitoring system. Section 4 details analyses Data Pre-Processing. Section 5 describes the experiments’ setup, as well as the results of our practical evaluations. Finally, Section 6 concludes this paper.

2. Literature Review
This section discusses the literature review analysis of some papers in relation to mobile application for diabetic patient monitoring using machine-learning algorithms. A description of the proposed systems and the used algorithms in this work are given. In their work, the authors presented an intelligent architecture for the surveillance of diabetic disease that monitor the health of diabetic patients through
sensors integrated into smartphones [7]. In another work, the prediction of the diabetes types using analysis algorithms and Hadoop map-reduce, prediction of complications, and the prediction of the type of treatment have been investigated [8], while in [9], the authors proposed a system that can perform predictions for Mellitus which is a type of diabetes using Hadoop/MapReduce. A new system for the prediction of the glucose concentration has been proposed in [10]. Where the data generated by the Continuous Glucose Monitoring can be analyzed by the glucoSim software using the Kalman Filter (KF) to reduce the noise.

Regarding Many researchers have developed and implemented various analysis and prediction models using different data mining techniques. In [11], the authors used a classification technique with Naive Bayes and Decision Tree algorithms by using the Weka tool to find out patterns from the diabetes data sets.

In [12] authors use Naïve Bayes and Decision Trees in one model classification technique to study hidden patterns in the diabetes dataset. Authors in [13] used the C4.5 decision tree algorithm, Neural Network algorithm, K-means clustering algorithm, and Visualization to predict diabetes for patients.

In this paper, we are using four machine learning algorithms, Naïve Bayes, Random Forest, OneR, and J48 classification algorithm from the diabetes data set, to test the most powerful algorithm to determine the patient’s level of risk. No one of the revised systems is specially designed and developed for diabetic patient monitoring. This work develops a new application using IoT for monitoring diabetic patients via Wi-Fi, but we are not limited to the monitoring, but we added the application of machine learning algorithms to classify the data generated by sensor and recovered by our application to classify and analyze it.

3. **Diabetic Patient Monitoring System**

In this part, we try to show our proposed system. It is a diabetic patient monitoring system based on node MCU. This system consists of three parts, the sensor part for data collection, the data analysis part, and the processing of the collected data. The diabetic patient monitoring system measures glucose level for the diabetic patients using glucose sensors and the data captured transmitted using Node MCU with the help of the IoT platform to the database for storage and treatment. The measured data are classified and analyzed by machine learning algorithms. The result of the classification is sent to the doctor to check the measured value, while a message is sent to the patients in an emergency case.

3.1. **General description:**

The general description of our proposed system for monitoring diabetic patients is presented in Fig. 1.
3.1.1. Glucose Sensor
A glucose sensor is a sensor intended for measuring the glucose level in a patient's blood. It has the same function as a continuous glucose monitoring system (CGM) and which can be either a sensor external placed on the skin or inserted under the skin.

3.1.2. Node MCU
Node MCU is a hardware and software IoT platform that integrates a 32-bit microcontroller. Composed of hardware in the form of an ESP8266 system on an ESP8266 chip manufactured by Expressive System. It is equipped with a Wi-Fi module, and which interfaces with the system using the USB cable for loading the program into the laptop.

3.1.3. Cloud Storage
The data collected by the glucose sensor is sent by the Node MCU and stored in the cloud and displayed on the mobile app. The data is also analyzed by machine learning algorithms and processed in the server. It is important to keep patient information as it may be used in the future. This database helps the doctor to interpret and interact in an emergency, in order to give the best and fastest diagnosis.

In this study, we develop a monitoring system for diabetic patients by using Node MCU to help diabetic patients manage their chronic illnesses themselves. The proposed system records glucose levels in diabetic patients and sends them automatically to the mobile app to save them on the cloud. The server can parse and process the data received to classify it. It can be considered as a platform for information on the level of glucose that allows interaction between patients and doctors. Using the learning machine algorithms installed in the server, we can automatically help users monitor their glucose levels and predict future changes in health. Fig. 2 shows the proposed system for diabetic patient monitoring.
Figure 2. Proposed System

4. Data Pre-processing
Due to the serious impact of diabetes on human health, it is relevant to predict diabetes early to better manage the disease. Artificial intelligence and machine learning algorithms are used widely for data classification and mainly for prediction. In our paper, we use machine learning to predict the state of diabetes in diabetic patients. Using machine learning, one can only predict with the data collected from the glucometer whether the diabetic patient is in the normal state or not without making the direct medical diagnosis. Therefore, our proposed system uses a supervised classification machine learning approach that captures the data set collected by the glucose sensor and makes the final decision on diabetes. Fig. 3 shows the proposed model.

Figure 3. Proposed Methodology Flowchart
4.1. Dataset and Attributes
In this part, we try to give the description of the dataset used as input for the classifiers implemented using various algorithms. Classification of data is an important step in this work that aims to classify the data as normal or affected by diabetes. The data studied included the records of 62 diabetic patients (44 men and 18 women) for 67 days with an average of three measurements per day. In this study, we used 4 classification algorithms (Naive Bayes [14], Random Forest [15], OneR [16], and SMO (Sequential Minimal Optimization)) [17]. In this study, the glucose level data were classified using multiple classification algorithms in the same data-mining programs. The glucose level data set contains 5 attributes and 12612 glucose level records. Fig. 4 shows blood glucose monitoring in diabetic patients three times a day, and Table 1 shows the type of data and description of the attributes.

![Figure 4. Screenshot of a patient’s stored glucose level](image)

| N° | Attributes     | Data type |
|----|----------------|-----------|
| 1  | Sex            | Boolean   |
| 2  | Age            | Digital   |
| 3  | Day            | Digital   |
| 4  | Glucose level  | Digital   |
| 5  | BGL            | Boolean   |

4.2. Weka Tool Description
To test the four classification algorithms that we used, we worked with the WEKA software (Waikato Environment for Knowledge Analysis: it is an open-source data extraction program), it is a toolbox of very popular machine learning and data mining to conduct data-driven research, is a data mining tool that allows us to classify data, analyze and determine the prediction accuracy of different mining algorithms. In this study, glucose level data was classified using multiple classification algorithms in the same data mining programs.

5. Experimental Results
In this section, we treat the experimental results obtained after the simulation of diabetic data with the classification algorithms Naive Bayes, Support Vector Machine, Random Forest, and Simple CART. The purpose of this simulation is to evaluate the performance of the classification algorithms that we use and to recommend the best suitable algorithm for prediction. Using the confusion matrix, we evaluate the prediction results using various evaluation measures such as the accuracy, classification accuracy, sensitivity, and the F-Measure.

Confusion matrix: is a matrix composed of the true value of a positive rate (TP), the true value of a negative rate (TN), the false value of a positive rate (FP), and the false value of a negative rate (FN).

**Table 2. Confusion Matrix**

|                | Positive (1) | Negative (0) |
|----------------|--------------|--------------|
| **POSITIVE (1)** | TP           | FP           |
| **NEGATIVE (0)** | FN           | TN           |

Precision: It is the ratio of predicted positive instances and the total of all predicted positive instances. Table 3 represents the experimental classification accuracy results and the training time of different algorithms.

**Table 3. Correctly and Incorrectly Classified Instances**

| Algorithms   | Correctly Classified Instances (%) | Incorrectly Classified Instances (%) | Time (S) |
|--------------|-----------------------------------|-------------------------------------|----------|
| Naive Bayes  | 90.23                             | 12.03                               | 0.05     |
| Random Forest| 96.45                             | 0.97                                | 0.02     |
| OneR         | 55.17                             | 60.33                               | 0.03     |
| SMO          | 7.14                              | 5.09                                | 0.03     |

Fig. 5 shows the classification accuracy value of all classifiers.
Figure 5. Accuracy value of all algorithms.

Fig. 6 shows the training time of all classifiers.

Figure 6. Time taken by all algorithms

From the following equations, we give the sensitivity values, precision, specificity, recall, and F-Measure for all the algorithms. Table. 4 represents the sensitivity, Precision, Specificity, Recall, and F-Measure Value of all classifiers.

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (2)
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \quad (3)
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \quad (4)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (5)
\]

\[
\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)
\]

Table 4. Values of Sensitivity, Specificity, Precision, Recall, and F-measure for algorithms.

| Algorithms     | Precision | Recall | F-measure |
|----------------|-----------|--------|-----------|
| Naive Bayes    | 0.91      | 0.85   | 0.83      |
| Random Forest  | 0.56      | 0.67   | 0.56      |
| OneR           | 0.78      | 0.58   | 0.91      |
| SMO            | 0.89      | 0.92   | 0.77      |

Fig. 7 shows the performance results of each classifier in terms of Precision, Recall, and F-measure.
Figure 7. Graphical representation of Precision, Recall, and F-measure.

Table 5 represents the sensitivity, Specificity, and Accuracy value of all classifiers.

**Table 5. Value of Sensitivity, Specificity, and Accuracy**

| Algorithms   | Sensitivity | Specificity | Accuracy |
|--------------|-------------|-------------|----------|
| Naive Bayes  | 97.34%      | 50.34%      | 85.11%   |
| Random Forest| 98.56%      | 84.50%      | 96.05%   |
| OneR         | 92.95%      | 18.27%      | 70.40%   |
| SMO          | 81.30%      | 77.17%      | 89.62%   |

Figure 8 represent the performance results of Sensitivity, Specificity and Accuracy for each classifier.
In this experimental study, four machine-learning algorithms were used Random Forest, OneR, SMO, and Naive Bayes. All these algorithms were applied to the diabetic patient dataset. Predicting accuracy is the main evaluation parameter that we used in this work. From Table 5, we can notice that the precision obtained by the Random Forest (98.56%) is better than the precision obtained by SMO, OneR, and Naive Bayes. It is also easy to see that the Random Forest has the highest value of the correctly classified instances and the lowest value of the incorrectly classified instances compared to other classifiers (see Fig. 8).

The accuracy of the algorithms was measured and presented in Fig. 8. OneR gives 70.40% accuracy, SMO gives 89.62% accuracy, 96.05% accuracy was achieved by Random Forest, and Naive Bayes gives 85.11%. Therefore, Random Forest achieved the highest accuracy, which is 96.05%. From the experimental results obtained. Therefore, the study leads us to conclude that the Random Forest algorithm achieves the best performance in terms of precision and accuracy.

6. Conclusion
In this study, several machine-learning algorithms are applied for classification on a data set. So, in this study, we have employed four main algorithms: OneR, Random Forest, SMO, and Naive Bayes on the diabetic datasets. These algorithms have been used for experimentation on WEKA tool to predict Diabetic patient data. We tried to compare the efficiency and the effectiveness of the cited algorithms in terms of accuracy, precision and sensitivity. The most objective is to choose the best classification accuracy. The overall performance of the Random Forest algorithm to predict diabetes disease is better than SMO, OneR, and Naive Bayes algorithms. In the future work will focus on the integration of other methods into the used model for tuning the parameters of models for better accuracy.

References

[1] G. J., Buxya. G.R, Marusic. S, &Palaniswami. M., "Internet of Things(IoT) A vision, architectural elements, and future directions". Future generation computer systeme, 1645-1660, 2013.
[2] Yuehong YIN, Y. Z," The internet of things in healthcare: An overview". Journal of Industrial Information Integration, 2016.
[3] L. Guariguata, D.R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, J.E. Shaw, "Global estimates of diabetes prevalence for 2013 and projections for 2035", Diabetes Res. Clin. Pract. vol.103, pp:137–149, 2014.
[4] P. K. Dhillon, S. Kalra, "Secure multi-factor remote user authentication scheme for Internet of Things environments", International Journal of Communication Systems, 2017.
[5] Rghioui, A.; Lloret, J.; Sendra, S.; Oumnad, A. Glucose Data Classification for Diabetic patient Monitoring. Applied Sciences. 2019, 9, 4459.
[6] Z. Mian, K. L.Hermayer, A. Jenkins, "Continuous Glucose Monitoring: Review of an Innovation in Diabetes Management", The American Journal of the Medical Sciences Vol. 358, pp: 332-339, Issue 5, November 2019.
[7] A. Rghioui, J. Lloret, M. Harrane, A. Oumnad, “A Smart Glucose Monitoring System For Diabetic Patient”, Electronics.9 (4), 678, 2020.

[8] [12] N. M. S. kumar, T. Eswari, P. Sampath, and S. Lavanya, “Predictive Methodology for Diabetic Data Analysis in Big Data,” Procedia Comput. Sci., vol. 50, pp. 203–208, Jan. 2015

[9] Ahmed, H.B.; Serener, A. Effects of External Factors in CGM Sensor Glucose Concentration Prediction. Procedia Comput. Sci. 2016, 102, 623–629

[10] K Saravananathan, Velmurugan T, "Analyzing Diabetic Data using Classification Algorithms in Data Mining", Indian Journal of Science and Technology 9(43), 2016.

[11] Dr Saravanan Kumar N M, Eswari T, Sampath P and Lavanya S,” Predictive Methodology for Diabetic Data Analysis in Big Data”, 2nd International Symposium on Big Data and Cloud Computing, 2015.

[12] Dost Muhammad Khan1, Nawaz Mohumudally2, “An Integration of K-means and Decision Tree (ID3) towards a more Efficient Data Mining Algorithm”, Journal Of Computing, Volume 3, Issue 12, December 2011.

[13] Luca Catarinucci, Danilo de Donno, Luca Mainetti, Luciano Tarricone, An IoT-Aware Architecture for Smart Healthcare Systems, IEEE Internet Things J. 2 (6) (2015) 515–526.

[14] S. Liver Disease Prediction Using Bayesian Classification, Special Issues, 4th National Conference on Advance Computing, Application Technologies, May 2014.

[15] B. D. Kanchan; M. M. Kishor. Study of machine learning algorithms for special disease prediction using principal of component analysis. International Conference on Global Trends in Signal Processing, Information Computing and Communication. 2016, 5. 5-10.

[16] O. N. Al Sayaydeha ; M. F. Mohammad, “Diagnosis of The Parkinson Disease Using Enhanced Fuzzy Min-Max Neural Network and OneR Attribute Evaluation Method”, International Conference on Advanced Science and Engineering (ICOASE), 2019.

[17] F. Wu ; D. Shao, "Study of SMO Parameter Optimization Based on PMSM Finite Element Simulation", IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2019.