Industrial Engineering Journal

ISSN: 0970-2555

Volume : 52, Issue 12, December : 2023

MACHINE LEARNING BASED DIABETES PREDICTION USING DECISION TREE J48

Aftab Nadaf, Student,

Prof. Sandip B. Shrote Professor

Electronics and Communication Engineering, School of Engineering and Science, MIT ADT University, Pune, India. [1]aftabn216@gmail.com, [2]sandip.shrote@mituniversity.edu.in

***Abstract*—Diabetes is a chronic disease with the potential to cause a worldwide health care crisis. According to International Diabetes Federation 382 million people are living with diabetes across the entire world. By 2035, this will be doubled as 592 million. Diabetes is a disease caused due to the increase level of blood glucose. This high blood glucose produces the symptoms of frequent urination, increased thirst, and increased hunger. Diabetes is a one of the leading causes of blindness, kidney failure, amputations, heart failure and stroke. When we eat, our body turns food into sugars, or glucose. At that point, our pancreas is supposed to release insulin. Insulin serves as a key to open our cells, to allow the glucose to enter and allow us to use the glucose for energy. But with diabetes, this system does not work. Type 1 and type 2 diabetes are the most common forms of the disease, but there are also other kinds, such as gestational diabetes, which occurs during pregnancy, as well as other forms. Machine learning is an emerging scientific field in data science dealing with the ways in which machines learn from experience. The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by Decision tree algorithm.**

***Index Terms*—Machine Learning, Diabetes, Decision tree, Accuracy.**

I. Introduction:

Diabetes is the fast-growing disease among the people even among the youngsters. In understanding diabetes and how it develops, we need to understand what happens in the body without diabetes. Sugar (glucose) comes from the foods that we eat, specifically carbohydrate foods. Carbohydrate foods provide our body with its main energy source everybody, even those people with diabetes, needs carbohydrate. Carbohydrate foods include bread, cereal, pasta, rice, fruit, dairy products, and vegetables (especially starchy vegetables). When we eat these foods, the body breaks them down into glucose. The glucose moves around the body in the bloodstream. Some of the glucose is taken to our brain to help us think clearly and function. The remainder of the glucose is taken to the cells of our body for energy and to our liver, where it is stored as energy that is used later by the body. For the body to use glucose for energy, insulin is required. Insulin is a hormone that is produced by the beta cells in the pancreas. Insulin works like a key to a door. Insulin attaches itself to doors on the cell, opening the door to allow glucose to move from the blood stream, through the door, and into the cell. If the pancreas is not able to produce enough insulin (insulin deficiency) or if the body cannot use the insulin it produces (insulin resistance), glucose builds up in the bloodstream (hyperglycaemia) and diabetes develops. Diabetes Mellitus means elevated levels of sugar (glucose) in the blood stream and in the urine. Types of Diabetes Type 1 diabetes means that the immune system is compromised, and the cells fail to produce insulin in sufficient amounts. There are no eloquent studies that prove the causes of type 1 diabetes and there are currently no known methods of prevention. Type 2 diabetes means that the cells produce a low quantity of insulin, or the body can't use the insulin correctly. This is the most common type of diabetes, thus affecting 90% of persons diagnosed with diabetes. It is caused by both genetic factors and the manner of living. Gestational diabetes appears in pregnant women who suddenly develop high blood sugar. In two thirds of the cases, it will reappear during subsequent pregnancies. There is a great chance that type 1, or type 2 diabetes will occur after a pregnancy affected by gestational diabetes. Symptoms of Diabetes

http://doi.org/10.36893/IEJ.2023.V52I12.006-011

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

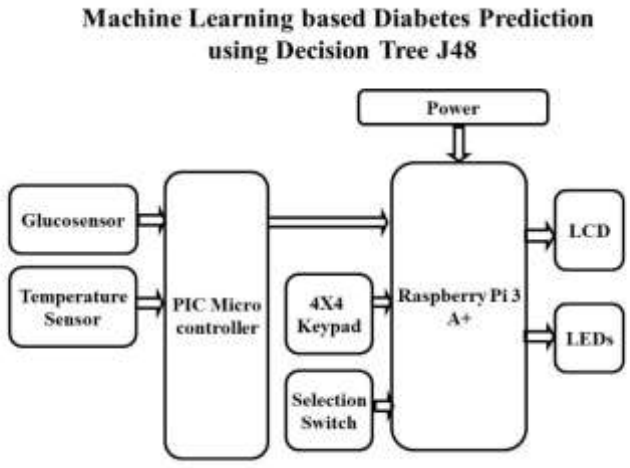
Frequent Urination, Increased thirst, Tired/Sleepiness, Weight loss, Blurred vision, Mood swings, Confusion and difficulty concentrating, frequent infections Causes of Diabetes Genetic factors are the main cause of diabetes. It is caused by at least two mutant genes in the chromosome 6, the chromosome that affects the response of the body to various antigens. Viral infection may also influence the occurrence of type 1 and type 2 diabetes. Studies have shown that infection with viruses such as rubella, Coxsackievirus, mumps, hepatitis B virus, and cytomegalovirus increase the risk of developing diabetes.

II. LITERATURE REVIEW:

1.Yasodhaet al. [1] uses the classification on diverse types of datasets that can be accomplished to decide if a person is diabetic or not. The diabetic patient's data set is established by gathering data from hospital warehouse which contains two hundred instances with nine attributes. These instances of this dataset are referring to two groups i.e., blood tests and urine tests. In this study the implementation can be done by using WEKA to classify the data and the data is assessed by means of 10-fold cross validation approach, as it performs very well on small datasets, and the outcomes are compared. The naïve Bayes, J48, REP Tree and Random Tree are used. It was concluded that J48 works best showing an accuracy of 60.2% among others

2. Aiswaryaet al. [2] aims to discover solutions to detect the diabetes by investigating and examining the patterns originate in the data via classification analysis by using Decision Tree and Naïve Bayes algorithms. The research hopes to propose a faster and more efficient method of identifying the disease that will help in well-timed cure of the patients. Using PIMA dataset and cross validation approach the study concluded that J48 algorithm gives an accuracy rate of 74.8% while the naïve Bayes gives an accuracy of 79.5% by using 70:30 split. Gupta et al. [3] aims to find and calculate the accuracy, sensitivity and specificity percentage of numerous classification methods and tried to compare and analyse the results of several classification methods in WEKA, the study compares the performance of same classifiers when implemented on some other tools which includes Rapidminer and Matlabusing the same parameters (i.e., accuracy, sensitivity and specificity). They applied JRIP, Jgraft and BayesNet algorithms. The result shows that Jgraft shows highest accuracy i.e 81.3%, sensitivity is 59.7% and specificity is 81.4%. It was also concluded that WEKA works best than Matlab and Rapidminner. Lee et al. [4] focus on applying a decision tree algorithm named as CART on the diabetes dataset after applying the resample filter over the data. The author emphasis on the class imbalance problem and the need to handle this problem before applying any algorithm to achieve better accuracy rates. The class imbalance is a mostly occur in a dataset having dichotomous values, which means that the class variable has two possible outcomes and can be handled easily if observed earlier in data preprocessing stage and will help in boosting the accuracy of the predictive model.

III. HARDWARE IMPLEMENTATION:



3.1 Block diagram of Machine Learning based Diabetes Prediction using Decision Tree J48 Algorithm

The main blocks of this project are: Glucosensor, Temperature sensor, PIC Microcontroller. 4X4 Keypad, Selection switches, Raspberry Pi, and LCD. This project makes use of a Raspberry Pi, which is programmed, with

http://doi.org/10.36893/IEJ.2023.V52I12.006-011

the help of Python. This Raspberry Pi can communicate with input and output modules. Sensors like glucosensor and temperature sensor give parameters such as blood glucose levels and body temperature levels whose values are given to PIC microcontroller which in turn gives the output to a Raspberry Pi which acts as the main controlling part of this project. Using selection switches, the user can select either the automatic mode or the manual mode. Through 4X4 keypad, the user can enter dataset values which are used for decision tree J48 algorithm. According to given values and the sensors data given by the microcontroller, the Raspberry Pi predicts the chances of diabetes of a person and displays it on LCD.

IV. Related Work:

The brief introduction of different modules used in this project is discussed below:

4.1. PIC16f72 Microcontroller:

Microcontroller:

The PIC16F73-I/SP is a 8-bit Flash-based CMOS Microcontroller. The PIC16F73 features 5 channels of 8-bit Analogue-to-digital (A/D) converter with 2 additional timers, 2 capture/compare/PWM functions and the synchronous serial port can be configured as either 3-wire Serial Peripheral Interface (SPI™) or the 2-wire Inter-Integrated Circuit (I²C™) bus and a Universal Asynchronous Receiver Transmitter (USART). The flash program memory is readable during normal operation over the entire VDD range. It is indirectly addressed through Special Function Registers (SFR).

- High performance RISC CPU
- All single-cycle instructions except for program branches which are two cycles
- Power-on reset (POR)
- Power-up timer (PWRT) and oscillator start-up timer (OST)
- Watchdog timer (WDT) with its own on-chip RC oscillator for reliable operation
- Programmable code protection
- Power saving sleep mode
- In-Circuit Serial Programming (ICSP™) via two pins
- -40 to +85°C Temperature range (industrial)

4.2. 4X4 keypad:

These Keypad modules are made of thin, flexible membrane material. The 4 x4 keypad module consists of 16 keys, these Keys are organized in a matrix of rows and columns. All these switches are connected to each other with a conductive trace. Normally there is no connection between rows and columns. When we will press a key, then a row and a column make contact.

4.3 Raspberry Pi 3 A+:

The Raspberry Pi 3 Model A+ extends the Raspberry Pi 3 range into the A+ board format.

- Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz
- 512MB LPDDR2 SDRAM
- 2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN, Bluetooth 4.2/BLE
- Extended 40-pin GPIO header
- Full-size HDMI
- Single USB 2.0 ports
- Micro SD port for loading your operating system and storing data
- 5V/2.5A DC power input

4.4 LCD Display:

http://doi.org/10.36893/IEJ.2023.V52I12.006-011

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

The term LCD stands for liquid crystal display. It is one kind of electronic display module used in an extensive range of applications like various circuits & devices like mobile phones, calculators, computers, TV sets, etc. These displays are mainly preferred for multisegmented light-emitting diodes and seven segments. The main benefits of using this module are inexpensive; simply programmable, animations, and there are no limitations for displaying custom characters, special and even animations, etc.

4.5 Non-invasive glucose concentration measurement circuit:

The proposed work is based on NIR optical technique. NIR light source of 940 nm wavelength is chosen because it is suitable for measuring blood glucose concentration. The sensing unit consists of NIR emitter and NIR receiver (photodetector) positioned on either side of the measurement site (fingertip). When the NIR light is propagated through the fingertip in which it interacts with the glucose molecule, a part of NIR light gets absorbed depending on the glucose concentration of blood and remaining part is passed through the fingertip. The amount of NIR light passing through the fingertip depends on the amount of blood glucose concentration. The transmitted signal is detected by the photodetector. The output current of the photo detector is converted into voltage signal and then it is filtered and amplified.

4.6 Temperature Sensor:

The LM35 sensor series are preciseness integrated-circuit temperature sensors, whose output voltage is linearly proportional to the Celsius (Centigrade) temperature.

To detect the heat produced during fire transpiration we use temperature sensor.

The Temperature Sensor LM35 sensor series are preciseness integrated-circuit temperature sensors, whose output voltage is linearly proportional to the Celsius (Centigrade) temperature.

4.7 Decision Tree J48 Algorithm:

J48 is a machine learning decision tree classification algorithm based on Iterative Dichotomiser 3. It is very helpful in examine the data categorically and continuously.
The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. The ID3 algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking.

Decision Trees are flowchart-like tree structures of all the possible solutions to a decision, based on certain conditions. It is called a decision tree as it starts from a root and then branches off to several decisions just like a tree. The tree starts from the root node where the most important attribute is placed. The branches represent a part of entire decision, and each leaf node holds the outcome of the decision.

A decision tree is a straightforward description of the splits found by the algorithm. Each terminal (or “leaf”) node describes a particular subset of the training data, and each case in the training data belongs to exactly one terminal node in the tree.

Feature Importance in Decision Trees Feature importance rates how important each feature is for the decision a tree makes. It is a number between 0 and 1 for each feature, where 0 means “not used at all” and 1 means “perfectly predicts the target”. Feature “Glucose” is by far the most important feature.

http://doi.org/10.36893/IEJ.2023.V52I12.006-011

V. CONCLUSION:

Integrating features of all the hardware components used have been developed in it. Presence of every module has been reasoned out and placed carefully, thus contributing to the best working of the unit. Secondly, using highly advanced ICs with the help of growing technology, the project has been successfully implemented. Thus, the project has been successfully designed and tested. The project "Machine Learning based Diabetes Prediction" was designed which can perform early prediction of diabetes for a patient with a higher accuracy by Decision tree algorithm.

VI. FUTURE SCOPE

In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.

VII. ACKNOWLEDGEMENT

We would like to thank all the authors of different research papers referred during writing this paper. It was very knowledge gaining and helpful for the further research to be done in future.

VIII. RESULTS:

The paper presents the design of "Machine Learning Based Diabetes Prediction using Decision Tree J48". The main objective of this design is to predict diabetes in its early stage in the patients with the highest percent of accuracy.

REFERENCES:

[1] Sajida Perveena, Muhammad Shahbaza, Aziz Guergachib, Karim Keshavjeec, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes" Procedia Computer Science 82 (2106) 115 –121.

[2] Deepti Sisodia, Dilip Singh Sisodia, "Prediction of Diabetes using Classification Algorithms", International Conference on Computational Intelligence and Data Science (ICCIDS2018)

[3] SantiWulanPurnami, AbdullahEmbong, JasniMohdZainand S.P. Rahayu, "A New Smooth Support Vector Machine and Its Applications in Diabetes Disease Diagnosis", Journal of Computer Science 5 (12): 1003-1008, 2009

[4] Faezeh Ensan, Mohammad Hossien Yaghmaee, Ebrahim Bagheri, "FACT: A new Fuzzy Adaptive Clustering Technique", The 11th IEEE Symposium on Computers and Communications, Sardinia, 26-29 June2006 [5] Aishwarya, Gayathri, Jaisankar, "A Method for Classification Using Machine Learning Technique for Diabetes", International Journal of Engineering and Technology2013.

[6] Dhomse Kanchan B., M.K.M., 2016. Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis, in 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, IEEE

[7] Esposito, F., Malerba, D., Semeraro, G., Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE Transactions on Pattern Analysis and MachineIntelligence19,476–491.

[8] Fatima, M., Pasha, M., 2017.SurveyofMachineLearningAlgorithmsfor Disease Diagnostic. Journal of Intelligent Learning Systems and Applications.

http://doi.org/10.36893/IEJ.2023.V52I12.006-011

[9] Kumar, P.S., Umatejaswi, V., 2017. Diagnosing Diabetes using Data Mining Techniques. International Journal of Scientific and Research Publications7.

[10] A. Mary Posonia, V.L. Jyothi (2016)," Extraction of perfect protein sequences with minimal processing cost using enhanced B+ tree algorithm", Biomedical Research, special issue on S12345-S6789

[11] A. Mary Posonia, Dr. V.L. Jyothi (2015)," Improving Data Access Performance by Reverse Indexing", International Journal of engineering and Technology (IJET), Vol 7 No 3, pp-1057- 1061

[12] Mary Posonia, Dr. V. L. Jyothi, "XML Document Retrieval by Developing an Effective Indexing Technique", in IEEE International Conference on IcoAC, MIT, Chennai, 2014, IEEE, DO I: 10.1109/ICoAC.2014.7229758, ISSN - 2377-6927

[13] Vimal Kumar S., Vasudevan S. and Mary Posonia A, "Urban Mode of Dispatching Students from Hostel", ARPN Journal of Engineering and Applied Science ,2017, Vol.12, No. 13

http://doi.org/10.36893/IEJ.2023.V52I12.006-011