VOLUME - 12, ISSUE - 12, DECEMBER - 2023 • PRINT ISSN No. 2277 - 8160 • DOI : 10.36106/gjra

Original Research Paper

Engineering

A NOVEL APPROACH ON MACHINE LEARNING BASED DIABETES PREDICTION IN REAL TIME

Aftab Nadaf	Student.
Sandip B. Shrote	Professor.

ABSTRACT The Project 'Machine Learning based Diabetes Prediction Using Decision Tree J48'aims to an innovative side in healthcare technology. At its core lies the designed system that has the predictive analysis of machine learning, specifically employing the Decision Tree J48 algorithm. The hardware foundation is based upon the Raspberry Pi Model B+, functioning as the central part. Collaborating seamlessly, a PIC microcontroller facilitates analog-to-digital conversion, ensuring precise data from temperature and glucose sensors. The human-computer interaction is streamlined through a user-friendly interface, featuring a 4x4 keypad and an LCD display. On the software front, the project has of Microsoft Visual Studio Code as the primary Python development environment. The machine learning library empowered by the scikit-learn module, and the algorithmic implementation of the Decision Tree J48 for predictions. In this both sensor and data set modes has been implemented. Users are empowered to contribute diverse datasets, fostering a personalized dimension to the prediction process. This consists of hardware and software, including by the seamless interaction of sensors, microcontrollers, and machine learning algorithms, showcases the project's approach. Beyond a technological feat, this project addresses a critical need in healthcare early diabetes risk detection.

KEYWORDS : Machine Learning, Diabetes Prediction, Healthcare Analytics, Scikit-Learn.

INTRODUCTION

The project aims to transform healthcare by introducing an automated system for predicting diabetes risk using the Decision Tree J48 algorithm. The collaboration of hardware and software components is pivotal in enhancing early detection and intervention. The hardware ensemble includes the Raspberry Pi Model B+, a robust single-board computer with essential features, a PIC Microcontroller for analog-to-digital conversion, and vital sensors such as the LM35 temperature sensor and a glucose sensor. Additional components like an LCD display, 4x4 keypad, and selection switch constitute the user interface, providing information presentation and user control. On the software front, Microsoft Visual Studio Code serves as the versatile code editor for Python development, while the scikit-learn module implements the Decision Tree J48 algorithm for accurate diabetes risk prediction. The Raspberry Pi OS acts as the operating system, facilitating seamless execution of Python scripts and the entire diabetes prediction system. The harmonious integration of these hardware and software elements results in a sophisticated embedded system, empowering users with predictive insights for proactive healthcare management and disease prevention.

OBJECTIVE

Machine Learning Application:

Develop and integrate a machine learning model using the J48 Decision Tree Algorithm for accurate prediction of diabetes risk.

Implement the model on the ARM-11 processor to process patient data effectively and provide reliable predictions.

Algorithm Implementation:

Successfully integrate and implement the J48 Decision Tree Algorithm on the Raspberry Pi, optimizing it to handle diverse patient data and provide accurate predictions.

Sensor Integration:

Integrate temperature and blood glucose sensors into the system, overcoming the challenge of analog-to-digital conversion for seamless and accurate data acquisition.

User Input Interface:

Implement a user-friendly interface using a 4x4 keypad, allowing users to input records, navigate between automated and manual modes, and interact effortlessly with the system.

Mode Selection:

Provide a toggle mechanism for users to switch between sensor mode and data set modes, ensuring flexibility in data collection and prediction. Create an interactive dashboard to customize settings for better results according to user preferences.

LITERATURE SURVEY

"Pattern Recognition and Machine Learning" by Christopher M. Bishop (2006):

It introduces fundamental concepts such as Bayesian networks, neural networks, and support vector machines. The book emphasizes a probabilistic approach to machine learning, providing a solid theoretical foundation. Practical examples and case studies are included to illustrate the application of machine learning algorithms.

"Machine Learning" by Tom M. Mitchell (1997):

It delves into the theoretical foundations of machine learning algorithms and their practical applications. The book emphasizes the importance of learning from data and adapting to new information. Case studies and examples help readers understand the real-world implications of machine learning.

"Induction of Decision Trees" by J. Ross Quinlan (1986):

It outlines the methodology for constructing decision trees from labeled training data. The paper discusses the use of entropy and information gain as measures for selecting the best attributes for splitting the data. Practical examples and experiments illustrate the effectiveness of decision trees in classification tasks.

"Machine Learning Techniques for Diabetes Prediction" by K. Srinivasan et.al. (2017):

It likely explores different algorithms, methodologies, and features used in the prediction process. The focus may include the use of relevant health data and variables in developing predictive models. The paper discusses the implications of machine learning in the context of diabetes management and healthcare.

METHODOLOGIES:

SOFTWARE METHODOLOGIES

The Development Environment:

- Utilization Microsoft Visual Studio Code for Python code development.

- Leverage of the features of VS Code, including syntax highlighting, code completion, and debugging tools.

Machine Learning Library:
- Employ the scikit-learn module as the machine learning library.
- Leverage scikit-learn's tools for data analysis and modeling, specifically focusing on the implementation of the J48 Decision Tree Algorithm.

Raspberry Pi Operating System:
- Install the Raspberry Pi OS, a Debian-based Linux distribution optimized for Raspberry Pi hardware.
- Choose the appropriate version of the OS, such as Raspbian Buster, to ensure compatibility with the Raspberry Pi Model B+.

Algorithm Implementation:
- Develop and integrate the J48 Decision Tree Algorithm using scikit-learn.
- Implement decision trees for accurate diabetes risk prediction based on the features derived from temperature and glucose sensors.

User Interface:
- Design an intuitive user interface using a 4x4 keypad for user inputs.
- Implement an LCD display to provide feedback on predictions or system status.

Data Set Handling:
- Implement functionalities to handle dynamic data collection and data set mode.
- Develop mechanisms to read, compare, and analyze data sets stored in the Raspberry Pi's memory.

Mode Selection:
- Provide a toggle mechanism for users to switch between sensor and dataset modes.
- Implement interactive dashboards to allow users to adjust settings for better predictions.

HARDWARE METHODOLOGIES

Raspberry Pi Integration:
- Establish a communication link between the Raspberry Pi and the PIC microcontroller.
- Utilize GPIO pins for serial communication and information exchange.

Pic Microcontroller (8-bit):
- Interface the PIC microcontroller with temperature and glucose sensors for data acquisition.
- Implement ADC capabilities for converting analog sensor data to digital form.

Sensor Connections:
- Connect the analog output of the temperature sensor to a suitable analog input pin on the PIC microcontroller.
- Connect the analog output of the glucose sensor to another analog input pin on the PIC microcontroller.

Keypad And LCD Connections:
- Connect the 4x4 keypad to GPIO pins on the PIC microcontroller.
- Connect the LCD display to the Raspberry Pi using GPIO pins or any supported display interface (e.g., SPI or I2C).

Power Supply:
- Provide a stable power supply for sensors, PIC microcontroller, and supporting components.
- Ensure compatibility of voltage levels between the Raspberry Pi and the microcontroller.

Serial Communication:
- Establish a reliable serial communication link between the Raspberry Pi and the PIC microcontroller.
- Utilize UART for seamless data transfer between the components.

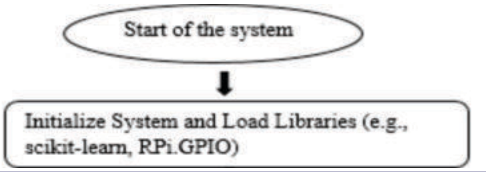
Data Range

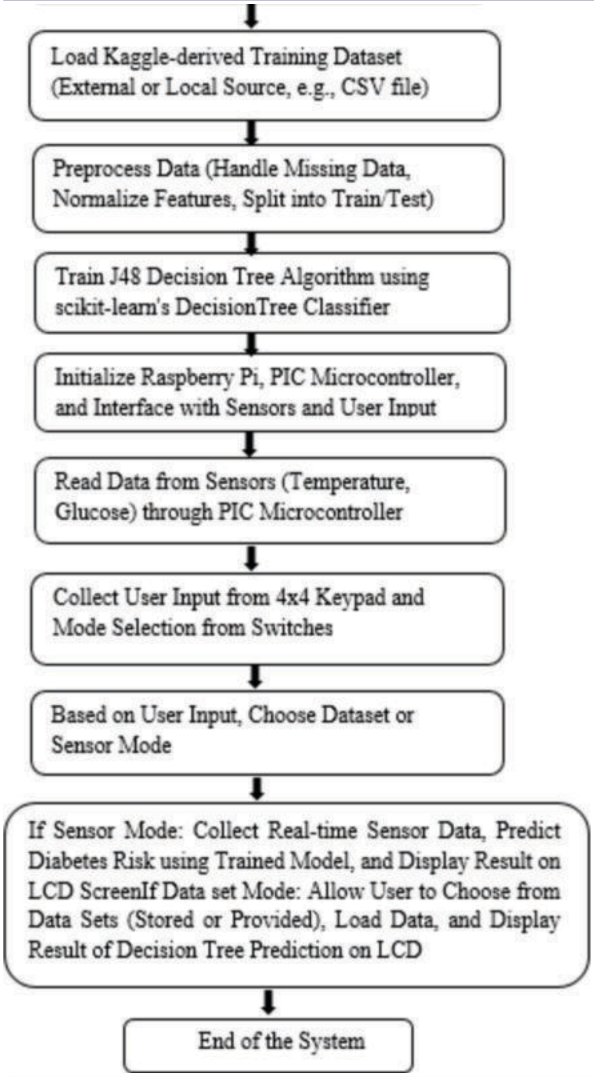
Defined six distinct types of outcomes based on ranges of glucose levels. Each outcome corresponds to a specific range, and these ranges are used to categorize individuals into different health states. Here's an elaboration on each outcome and its associated glucose level range in Milligrams per decilitre (mg/dL):
- **Excellent (0-90):** Individuals falling within the glucose level range of 0 to 90 mg/dL are classified as "Excellent." This range typically represents a healthy blood sugar level.
- **Normal (90-120):** The "Normal" category encompasses individuals with glucose levels ranging from 90 to 120 mg/dL. This range is considered normal and indicates a stable blood sugar level.
- **Borderline (120-130):** Individuals with glucose levels in the range of 120 to 130 mg/dL are classified as "Borderline." This range suggests a potential deviation from normalcy, warranting attention.
- **Prediabetes (130-150):** The "Prediabetes" category includes individuals with glucose levels ranging from 130 to 150 mg/dL. This range signifies an elevated blood sugar level that may precede the onset of diabetes.
- **Diabetes (150-180):** Individuals with glucose levels in the range of 150 to 180 mg/dL are categorized as having "Diabetes." This range indicates a higher and potentially unhealthy blood sugar level requiring medical attention.
- **Dangerous (180 and above):** The "Dangerous" category includes individuals with glucose levels equal to or exceeding 180 mg/dL. This range represents a critical condition, and individuals falling into this category may be at risk of severe health complications. Urgent medical intervention is typically required4.3 Raspberry Pi 3 B+:

DECISION TREE J48

- **Node Splitting:** J48 employs a top-down, recursive approach. It starts with the entire dataset and selects the feature that best splits the data into subsets. This process continues recursively for each subset until a stopping criterion is met.
- **Attribute Selection:** At each node, the algorithm decides which attribute (feature) to use for splitting. It selects the attribute that maximizes the information gain or minimizes the impurity of the subsets, ensuring the resulting branches are as pure as possible.
- **Stopping Criteria:** The recursive splitting continues until a predefined stopping condition is met. This could be reaching a certain depth of the tree, achieving a minimum number of instances per leaf, or other criteria to prevent overfitting.
- **Leaf Node Prediction:** Once the tree is constructed, new instances traverse the tree from the root to a leaf node. The majority class in the leaf node is then assigned as the predicted outcome.
- **Application in the Diabetes Prediction System**: In this project, the Decision Tree J48 algorithm is used for predicting the risk of diabetes based on the input data, which includes glucose levels and temperature.

Flowchart of Working Model





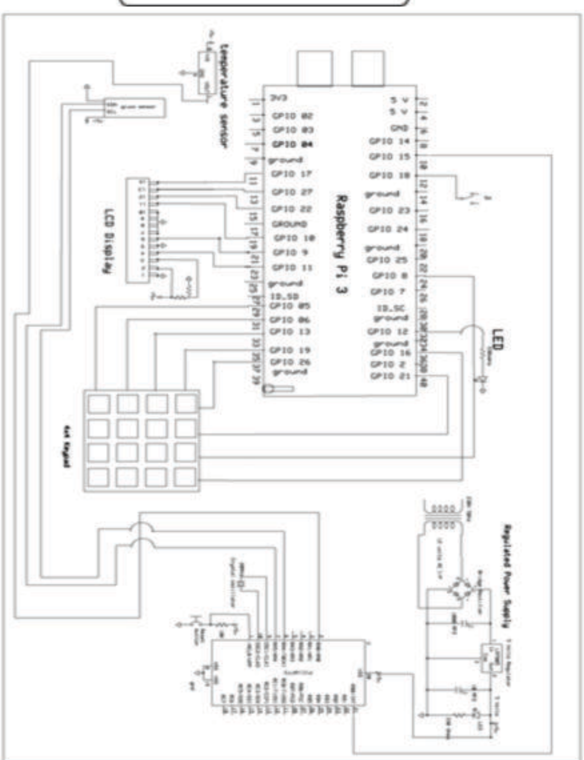


Fig. 4.6.1-Simlation Circuit Diagram.

RESULTS

Experiment 1

Table 5.1- Data Of The Patient

Outcome	Glucose Level (mg/dL)	Temperature Level (°F)
Excelllent	85	97.7
Normal	100	98.6
Borderline	125	98.2
Prediabetes	140	98.96
Diabetes	160	98.42
Dangerous	190	99.5

Plotted Graph

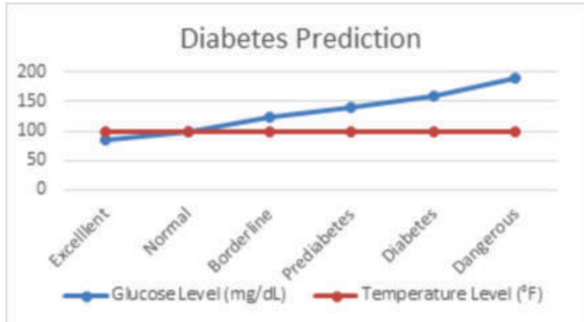


Fig 5.1 – Graph for Exp-1

In the provided patient data set glucose levels are paired with corresponding temperature levels, and each entry is associated with a predicted outcome. Glucose levels range from 85 to 190 mg/dL, and temperature levels are given in Fahrenheit, ranging from 97.7°F to 99.5°F. The outcomes are categorized into six types based on these levels. A glucose level of 85 mg/dL and a temperature of 97.7°F are labeled as "Excellent," indicating a healthy state. As glucose levels increase, the outcomes progress through "Normal," "Borderline," "Prediabetes," and "Diabetes." The highest glucose level of 190 mg/dL paired with a temperature of 99.5°F is categorized as "Dangerous," reflecting a critical health condition. This dataset is designed to simulate different health scenarios for the purpose of training and testing a diabetes prediction model.

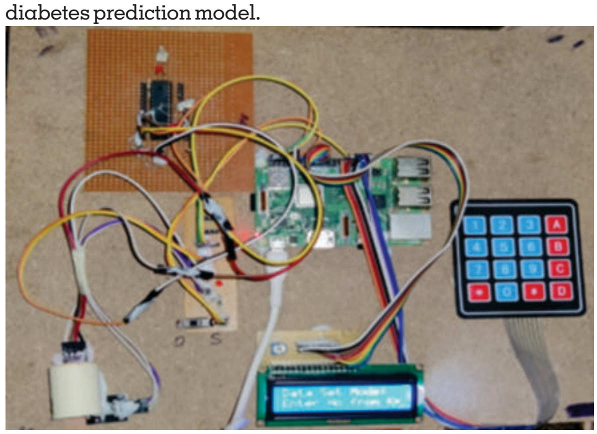


Fig. 5.2 Overall Hardware Kit

Fig. 5.3 Body temperature and glucose level of the user

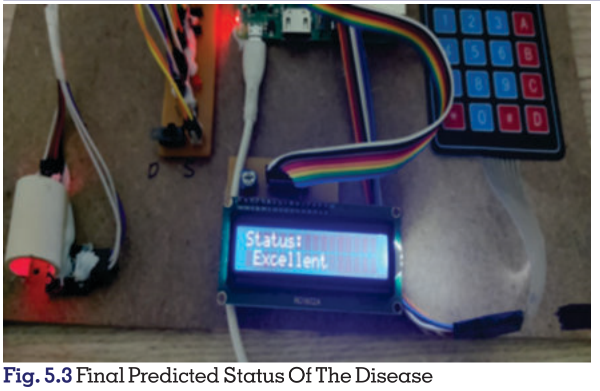


Fig. 5.3 Final Predicted Status Of The Disease

Future Scope

The project opens avenues for future enhancements and expansions in several directions like

- AlgorithmicImprovements-Exploration of advanced machine learning algorithms and ensemble techniques could enhance the prediction accuracy and robustness of the system.
- RemoteMonitoring-Implementing features for remote patient monitoring and data transmission could extend the system's applicability in telemedicine and healthcare IoT.

Acknowledgement

I 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. I would like to specially thank my guide Mr. Prof.Sandip B. Shrote for his kind guidance and valuable research carried throughout this project work.

CONCLUSION

The project successfully integrates the Decision Tree J48 algorithm into a system designed for diabetes prediction. Utilizing data collected from glucose and temperature sensors, the algorithm, implemented on a Raspberry Pi, demonstrates effective learning and prediction capabilities. The system ensures user-friendly interaction through a keypad, allowing users to choose between sensor and dataset modes. With a focus on both hardware and software components, the project addresses the critical need for proactive healthcare solutions by providing accurate and timely predictions regarding diabetes risk. 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.