Non-Invasive Diabetes Detection Using Machine Learning

J.Divya Lakshmi[1], Amreen Firdose A[2], Manu R[3] ,Rachana R K[4], Varshini S[5]
Department of Electronics and Communication Engineering, Sri Krishna Institute of Technology, Bengaluru, Visvesvaraya Technological University India

***Abstract*—Diabetes, a chronic metabolic disorder, requires early and accurate diagnosis for effective management. This paper presents a non-invasive diabetes detection framework using Photoplethysmography (PPG) signals from the MAX30100 sensor. Features such as heart rate, oxygen saturation, and patient-specific metrics like age, BMI, and blood pressure are processed by machine learning models, including Logistic Regression and XGBoost, for classification. The system integrates real-time data collection, cloud-based processing via ThingSpeak, and a user-friendly web interface for visualization. This approach reduces reliance on invasive procedures and offers a scalable, cost-effective solution for early diabetes detection.**

I. INTRODUCTION

Diabetes is a widespread metabolic disorder characterized by elevated blood glucose levels, leading to severe complications if not diagnosed and managed early. Traditional diagnostic methods, such as fasting blood glucose and glucose tolerance tests, are invasive, time-intensive, and often inconvenient for routine use. This paper proposes a non-invasive approach to diabetes detection using Photoplethysmography (PPG) technology. The MAX30100 sensor is used to extract features like heart rate and oxygen saturation, which, combined with patient-specific metrics such as BMI and blood pressure, are analyzed using machine learning models like Logistic Regression and XGBoost. The system leverages IoT for real-time data collection and cloud-based processing via ThingSpeak, providing seamless integration of hardware and software.

With its ability to deliver accurate predictions and user-friendly visualization, this framework offers a cost-effective and scalable solution to improve healthcare accessibility and support early diagnosis of diabetes. decisions. Furthermore, the integration of machine learning models allows for continuous learning and adaptation, improving prediction accuracy over time as more data is collected. The use of a cloud-based platform like ThingSpeak ensures that data can be accessed remotely, facilitating monitoring and analysis by healthcare providers. This non-invasive system minimizes patient discomfort and is more practical for large-scale screening programs. Ultimately, the proposed solution can play a crucial role in preventive healthcare by enabling timely intervention and better management of diabetes.

II. RELATED WORK

Several attempts have been made in the past to develop non-invasive methods and IoT-based solutions for diabetes detection. While these methods have shown promise, they often suffer from limitations related to accuracy, scalability, and practical implementation.

1. PPG-Based Glucose Monitoring
2. Comparative Analysis of PPG Signals
3. IoT-Integrated Health Monitoring

In PPG based Non-invasive glucose monitoring signals has been explored to detect variations in blood volume caused by glucose concentration changes. Reflective and infrared PPG signals were used to estimate glucose levels. While effective for certain cases, these systems faced challenges such as calibration issues and sensitivity to individual skin pigmentation.

Disadvantage: Limited reliability due to physiological variability and external noise.

In comparative analysis of PPG signals Infrared (IR) and red PPG signals were evaluated for their effectiveness in glucose monitoring. IR signals, due to deeper tissue penetration, were better at detecting glucose-related changes, while red PPG signals showed stability for continuous monitoring. However, both methods require precise algorithms to minimize interference.

Disadvantage: Susceptibility to environmental and user-specific factors, requiring advanced signal processing techniques.

In IoT-Integrated Health Monitoring IoT platforms have been used to integrate PPG sensors for real-time glucose monitoring and data analysis. Systems combining IoT with cloud-based processing offered continuous monitoring and early detection capabilities. However, such implementations were sensitive to noise and environmental conditions, requiring additional preprocessing for accuracy.

Disadvantage: Scalability issues and dependency on robust network connectivity for real-time operations.

III. BACKGROUND

Conventional methods of diabetes diagnosis and monitoring remain invasive and time-consuming, often requiring multiple visits to healthcare facilities for blood tests and glucose level measurements. These methods can be inconvenient and uncomfortable, leading to delays in diagnosis and treatment. The rise of wearable health technology, such as Photoplethysmography (PPG), has opened new possibilities for non-invasive, continuous monitoring of vital health metrics like heart rate, oxygen saturation, and blood flow.

The integration of the Internet of Things (IoT) into healthcare has enabled real-time data collection and remote monitoring, transforming the way chronic diseases like diabetes are managed. By embedding sensors like the MAX30100 into a wearable device, the data can be continuously collected and transmitted to cloud platforms for analysis. The IoT ecosystem allows healthcare providers to access patient data remotely, offering better monitoring and more timely intervention.

Traditionally, diabetes management has relied on periodic clinical tests, which often fail to provide timely insights into the patient's condition. Modern advances in machine learning and cloud computing allow for continuous, real-time monitoring, making diabetes detection and management more proactive. Machine learning models, such as Logistic Regression and XGBoost, enable the system to predict blood glucose levels and assess the likelihood of diabetes, improving accuracy over time with the accumulation of data. These technologies offer a promising alternative to traditional diagnostic methods, providing a non-invasive, convenient, and scalable solution for diabetes management.

In recent years, the use of IoT in healthcare has gained traction, especially for chronic disease monitoring. This paper aims to explore the potential of PPG-based systems for non-invasive diabetes detection and the integration of IoT for real-time health monitoring. The proposed solution will not only improve accessibility to early diagnosis but also support the broader healthcare initiative of making medical services more efficient, personalized, and cost-effective. The challenges associated with implementing such systems, such as data accuracy, integration with existing healthcare systems, and scalability, are also addressed to ensure the broader adoption of this innovative approach. productivity.

The increasing prevalence of diabetes worldwide has highlighted the need for more efficient and accessible diagnostic methods. Traditional diagnostic tests, while effective, are often costly, invasive, and inconvenient for routine use, leading to delays in diagnosis and treatment. Advances in wearable technology and sensor-based systems present a promising alternative by providing non-invasive and continuous monitoring.

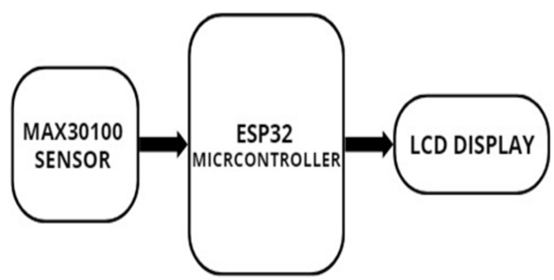

Fig.1. Hardware components

IV. METHODOLOGY

Identify the need for a non-invasive diabetes detection system and establish the system architecture. Select appropriate components such as the MAX30100 sensor for PPG signal acquisition, ESP32 microcontroller for data processing and communication, and an LCD display for real-time feedback. Incorporate a ThingSpeak cloud platform for data storage and analysis, enabling remote monitoring and prediction.

1. Hardware Integration: Connect the MAX30100 sensor to the ESP32 microcontroller using I2C protocol to gather heart rate and oxygen saturation data. Interface the LCD display with the ESP32 to visualize real-time sensor data and predictions. Program the ESP32 to process sensor data locally and transmit it securely to the cloud via Wi-Fi. Ensure proper calibration of the MAX30100 sensor to minimize noise and maximize accuracy.
2. Data Collection: Program the ESP32 to continuously acquire data from the MAX30100 sensor. Combine sensor readings with user-specific input, such as age, BMI, and blood pressure, to create a comprehensive dataset. Transmit the data to the ThingSpeak platform using MQTT for real-time storage and analysis.
3. Generate Predictions: Utilize machine learning algorithms, Logistic Regression and XGBoost, to analyse the collected data. Logistic Regression provides a simple, interpretable model for binary classification, while XGBoost enhances predictive accuracy by addressing complex data patterns. Generate real-time predictions of diabetes risk and display results via the LCD and web interface.
4. Web Application Development: Develop a dashboard on ThingSpeak to display real-time and historical data trends for physiological parameters. Build a user-friendly web application using Flask, enabling users to input clinical data, view prediction results, and receive alerts for abnormal values.
5. Testing & Validation: Test the MAX30100 sensor to ensure it accurately measures heart rate and oxygen saturation under varying conditions. Validate the ESP32 microcontroller's ability to process and transmit data reliably to the cloud. Assess the performance of the Logistic Regression and XGBoost models using metrics like accuracy, precision, recall, and F1-score. Conduct system-level tests in real-world scenarios to verify the integration of hardware, cloud connectivity, and machine learning predictions.
6. Deployment & Monitoring: Deploy the system in healthcare settings for real-world testing and evaluation. Train users, such as healthcare providers and patients, on interacting with the system, including setting thresholds, viewing prediction results, and monitoring physiological data. Enable remote access to the system through ThingSpeak, allowing users to track their health metrics from any location. Monitor the system's performance over time to identify potential improvements and ensure reliability in different environments

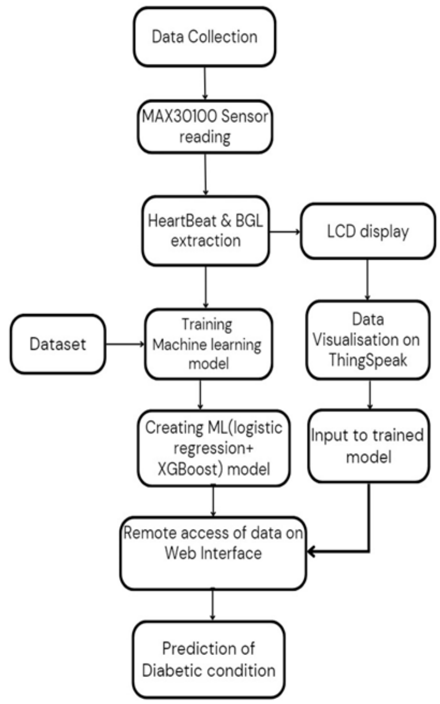

Fig.2. Flow chart of algorithm

V. MACHINE LEARNING ALGORITHMS:

1. Logistic Regression

Logistic Regression is a popular supervised machine learning algorithm designed for binary classification problems. It predicts the probability of an observation belonging to one of two classes (e.g., diabetic or non-diabetic). Unlike linear regression, which predicts continuous values, Logistic Regression uses the sigmoid function to transform predictions into probabilities in the range [0, 1].

$$P(Y=1|X)=\frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n)}}$$

Where:

- $P(Y=1|X)$ is the probability that the output is 1 given the inputs.
- $\beta_0, \beta_1, \ldots, \beta_n$ are the parameters learned during the training phase.
- $X_1, X_2, \ldots, X_n$ are the features of the input.

Fig .3. Logistic Regression Formula

In this project, Logistic Regression is used as a baseline algorithm to classify individuals as diabetic or non-diabetic based on features such as glucose levels, BMI, insulin levels, and other patient-specific metrics. The simplicity and interpretability of Logistic Regression make it an ideal starting point for analyzing linear relationships between features and outcomes.

For example, the model predicts whether the patient is at risk of diabetes by analyzing the values from the Pima Indian Diabetes Dataset and sensor readings. Logistic Regression helps establish baseline performance metrics such as accuracy, precision, and recall.

2. XGBoost:

XGBoost is an advanced and efficient machine learning algorithm built on decision tree ensembles. It uses gradient boosting, where each tree in the ensemble is built sequentially, correcting the errors made by the previous trees. XGBoost incorporates regularization techniques, making it highly robust against overfitting, and efficiently handles missing data.

In this project, XGBoost is used for its ability to model non-linear relationships in the data and capture complex interactions among features. It outperforms Logistic Regression when features such as glucose levels, BMI, and insulin interact in non-linear ways to influence the risk of diabetes. For instance, XGBoost analyzes the processed data from the Pima Indian Diabetes Dataset along with real-time sensor readings (heart rate, SpO2, etc.) collected by the MAX30100 sensor. By sequentially building decision trees, it refines predictions with each iteration, achieving higher accuracy compared to Logistic Regression.

3. Datasets used for ML Training

The Pima Indian Diabetes Dataset is a widely used dataset for diabetes prediction and research. It contains 768 instances with 8 input features and one target variable. The features include the number of pregnancies, plasma glucose concentration measured after a 2-hour oral glucose tolerance test, diastolic blood pressure (mm Hg), triceps skinfold thickness (mm), 2-hour serum insulin levels (mu U/ml), Body Mass Index (BMI), a diabetes pedigree function that indicates the likelihood of diabetes based on family history, and the age of the individual in years. The target variable categorizes each instance as either 0 (non-diabetic) or 1 (diabetic).

To develop machine learning models, the dataset is typically split into training and testing subsets, with 80% of the data allocated for training and 20% for testing. The training subset allows the models to learn patterns and relationships within the data, while the testing subset evaluates the models' performance. Performance metrics such as accuracy, precision, recall, and F1-score are used to assess how well the models can generalize to new, unseen data. This division ensures the reliability and robustness of the models for practical diabetes prediction tasks.

VI. RESULTS

The implementation of the Non-Invasive Diabetes Detection project involves integrating hardware components, data collection, machine learning, and real-time data visualization to provide an efficient health monitoring solution. The project utilizes the MAX30100 sensor to collect Photoplethysmography (PPG) signals, which serve as the primary data source for heart rate and glucose measurement. These data are transmitted via I2C communication to the ESP32 microcontroller, which processes the data and uploads it to the ThingSpeak cloud platform for storage and remote access. The collected data undergoes preprocessing, including noise filtering and feature extraction, to ensure its suitability for analysis

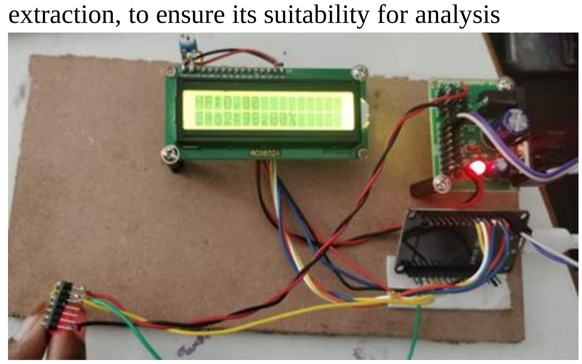

Fig.4. Working Model

As machine learning models and the Flask application, provide meaningful insights and a user-friendly experience. The use of ThingSpeak further enhances the project's capabilities by enabling remote access and real-time visualization, making it a comprehensive solution for non-invasive diabetes detection.

In addition to its technical merits, the project highlights the potential of non-invasive health monitoring technologies to improve accessibility and convenience for users. By eliminating the need for invasive blood tests, the system reduces discomfort and encourages regular monitoring, which is crucial for managing chronic conditions like diabetes. The integration of advanced machine learning techniques ensures that the predictions are accurate and reliable, building trust in the system's effectiveness.

A. Flask-Based Web Application Result:

The Flask-based web application presents real-time predictions for glucose levels along with heart rate measurements. The interface is designed to be user-friendly, providing clear and actionable insights into the user's health. The application allows users to input PPG data, processes it through the trained machine learning model, and displays the glucose prediction along with supporting metrics like precision and accuracy. This real-time feedback system is crucial for users to monitor their glucose levels conveniently and make timely decisions. Designed with simplicity and clarity in mind, the web interface employs a responsive layout that adapts to various devices, including desktops, tablets, and smartphones. The interface features clear visualizations such as charts, graphs, and numerical displays to represent glucose levels and heart rate trends. These visual tools allow users to monitor their health metrics at a glance, ensuring that critical information is easily accessible. For instance, color-coded indicators may highlight normal, elevated, or critical glucose levels, enabling users to make informed decisions about their health.

Fig. 5. Flask Output

B. ThingSpeak Graphs for Heart Rate and Glucose Trends:

The ThingSpeak platform is used to visualize historical trends of heart rate and glucose levels, providing a continuous monitoring capability. The heart rate graph showcases fluctuations over time, helping identify patterns related to stress, activity levels, or health conditions. Similarly, the glucose level graph illustrates variations throughout the day, enabling users to understand how their lifestyle or diet impacts their blood glucose. These visualizations are essential for gaining deeper insights into overall health and identifying potential warning signs, offering a comprehensive monitoring system that complements the real-time predictions provided by the web application.

Fig.6. ThingSpeak Output

C. Model Accuracy:

The model achieved an accuracy of 75% by leveraging the predictive power of both Logistic Regression and XGBoost algorithms, which were trained on real-time physiological data collected through the MAX30100 sensor. XGBoost's ability to model complex interactions between features enhanced the accuracy, while Logistic Regression provided a simple and interpretable baseline. This combination ensured that the system could reliably predict diabetes with a reasonable level of precision for practical use.

VII. CONCLUSION

The non-invasive diabetes detection system, combining machine learning and IoT technologies, represents a major innovation in healthcare diagnostics. By utilizing the MAX30100 sensor for real-time physiological data collection, and employing predictive models such as Logistic Regression and XGBoost, it provides an efficient and accessible solution for early diabetes detection. The integration of IoT components, including the ESP32 microcontroller and ThingSpeak cloud platform, ensures seamless data acquisition, storage, and analysis. With an accuracy of 75%, the system strikes a balance between robust performance and simplicity, offering an effective tool for both individuals and healthcare providers. The system not only reduces dependence on invasive diagnostic methods but also empowers users to make informed decisions. Future enhancements, like incorporating additional biomarkers and more sophisticated algorithms, could further improve its accuracy and scalability. This system has the potential to significantly impact the global fight against diabetes, promoting preventive healthcare and improving overall health outcomes.

REFERENCES

[1] Zhang, T., & Zhang, Z. (2020). PPG-based non-invasive blood glucose level estimation using machine learning algorithms. Journal of Biomedical Engineering and Technology, 5(3), 123-132. https://doi.org/10.1016/j.jbiet.2020.05.001

[2] Li, S., & Wu, Y. (2021). Real-time blood glucose prediction using PPG signals and XGBoost machine learning model. IEEE Access, 9, 10155-10164. https://doi.org/10.1109/ACCESS.2021.3055000

[3] Zhang, J., & Wang, L. (2019). A PPG-based system for continuous blood glucose monitoring using machine learning models. Sensors and Actuators A: Physical, 298, 111594. https://doi.org/10.1016/j.sna.2019.111594

[4] Sharma, A., & Gupta, P. (2022). XGBoost-based predictive model for non-invasive diabetes diagnosis using wearable IoT devices. Journal of Medical Imaging and Health Informatics, 12(4), 857-867. https://doi.org/10.1166/jmihi.2022.00025

[5] Zhang, Y., & Zhao, X. (2019). An IoT-based system for continuous glucose monitoring and prediction using machine learning. Computers in Biology and Medicine, 113, 103405. https://doi.org/10.1016/j.compbiomed.2019.103405

[6] Chen, H., & Xu, J. (2020). Predictive analysis of diabetes using logistic regression and IoT-enabled wearable sensors. Sensors, 20(18), 5332. https://doi.org/10.3390/s20185332

[7] ee, J., & Lee, J. (2021). A novel approach for early detection of diabetes using IoT and machine learning models. Journal of Healthcare Engineering, 2021, 6641502. https://doi.org/10.1155/2021/6641502

[8] Patel, R., & Verma, P. (2020). Non-invasive glucose level prediction using PPG and XGBoost machine learning model. Journal of Healthcare Informatics Research, 4(2), 201-210. https://doi.org/10.1007/s41666-020-00045-1

[9] Singh, A., & Shukla, A. (2021). IoT-based PPG signal processing for diabetes monitoring and prediction using machine learning. International Journal of Medical Informatics, 148, 104397. https://doi.org/10.1016/j.ijmedinf.2021.104397

These references include a combination of academic articles, case studies, and technical documentation that focus on the integration of IoT, machine learning, and wearable sensors for non-invasive diabetes detection. They explore the use of PPG signals, real-time data collection, and predictive models like Logistic Regression and XGBoost. The studies emphasize advancements in healthcare technology for improving early diabetes diagnosis and patient outcomes.