EARLY PREDICTION OF TYPE 2 DIABETES MELLITUS IN YOUNG ADULTS USING LSTM

Project Reference No.: *48S_BE_5823*

College : Reva University, Bengaluru
Branch : Electronics and Communication Engineering
Guide : Dr. Veena K N
Student(s): Ms. Amrutha HM
Mr. Chandan M

Keywords:

Type 2 Diabetes Mellitus (T2DM), Early Prediction, Long Short-Term Memory (LSTM), Machine Learning (ML), Internet of Things (IoT), Non-invasive Glucose Monitoring.

Introduction:

Type 2 Diabetes Mellitus (T2DM) is becoming increasingly prevalent among young adults, often going undiagnosed until serious complications arise [1]. The lack of early symptoms, combined with limited awareness and lifestyle-related factors, contributes to delayed detection and treatment [2]. Traditional glucose monitoring methods—mainly blood-based tests—are invasive, uncomfortable, and discourage regular health tracking, especially among youth who may not recognize the risk.

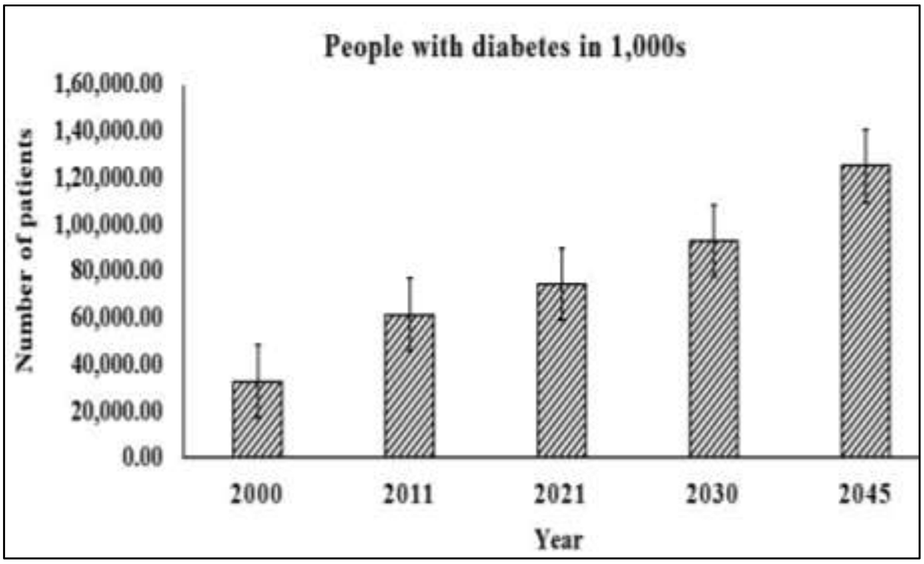

Figure 1: Diabetes statistics by IDF [3]

There is a growing demand for solutions that are non-invasive, easy to use, and capable of encouraging consistent health monitoring. With recent advancements in technology, the integration of Internet of Things (IoT) and Machine Learning (ML) has opened up new possibilities in the field of predictive and preventive healthcare. IoT-based sensors can collect real-time physiological and environmental data, while ML algorithms can process this data to identify hidden patterns, assess potential health risks, and deliver timely recommendations. Together, these technologies are enabling the development of smart, personalized, and accessible healthcare systems [4-7].

By integrating the insights from our LSTM prediction model [8] and the IoT sensor system, this project aspires to offer a cost-effective and accessible solution for monitoring diabetes risk, laying the groundwork for the next phase of our research. The combination of breath analysis and machine learning will enable real-time, non-invasive prediction of diabetes, which can be used for proactive health management and early intervention.

Objectives:

1.1 Early Prediction

This section focuses on identifying the real-world problem, understanding its impact on young adults, and highlighting the need for an effective and early diagnostic approach.

- Tackle Rising Health Concerns**:** Address the growing incidence of Type 2 Diabetes Mellitus (T2DM) in young adults.

- Identify Gaps in Diagnosis: Recognize the limitations of conventional diagnostic tools that often miss early symptoms.

- Prevent Health Complications**:** Emphasize the importance of timely diagnosis to reduce the risk of severe long-term effects.

Sensor - Based Monitoring

This part outlines how the project aims to solve the identified problem using technology—through a combination of machine learning and IoT—to build a smart, non-invasive, and real-time monitoring system.

- Non-Invasive Monitoring System**:** Design a prototype using IoT-based breath analysis for painless glucose monitoring.

- Apply Deep Learning Models **:** Utilize LSTM-based machine learning models to predict the risk of T2DM accurately.

- Integrate Real-Time Sensor Data **:** Use environmental and physiological data from sensors for more precise health analysis.

Methodology:

1.2 Early Prediction

The study focuses on developing a predictive model to assess the risk of Type 2 Diabetes Mellitus (T2DM) in young adults using machine learning techniques [8-9].

i. Medical and lifestyle data such as age, BMI, glucose, insulin, and other health metrics were collected from publicly available datasets.

ii. The data was cleaned and normalized to ensure consistency and accuracy.

iii. A deep learning model using Long Short-Term Memory (LSTM) networks was developed to detect patterns and predict diabetes risk.

iv. The model was trained and evaluated using performance metrics like accuracy, precision, recall, and F1-score, ensuring reliable predictions.

1.3 Sensor-Based Monitoring

This phase focuses on creating a non-invasive, real-time glucose monitoring system using breath samples, designed with IoT sensors and hardware [10].

i. MQ135 (gas sensor) and DHT11 (temperature & humidity sensor) were used and connected to the ESP32 microcontroller for collecting breath data.

ii. The system guides the user to take multiple breath samples to ensure consistent readings.

iii. Stability checks were applied to validate the consistency of the sensor data, ensuring accurate glucose estimation.

iv. The glucose levels were estimated based on the collected breath samples, but no further health recommendation was provided in this phase.

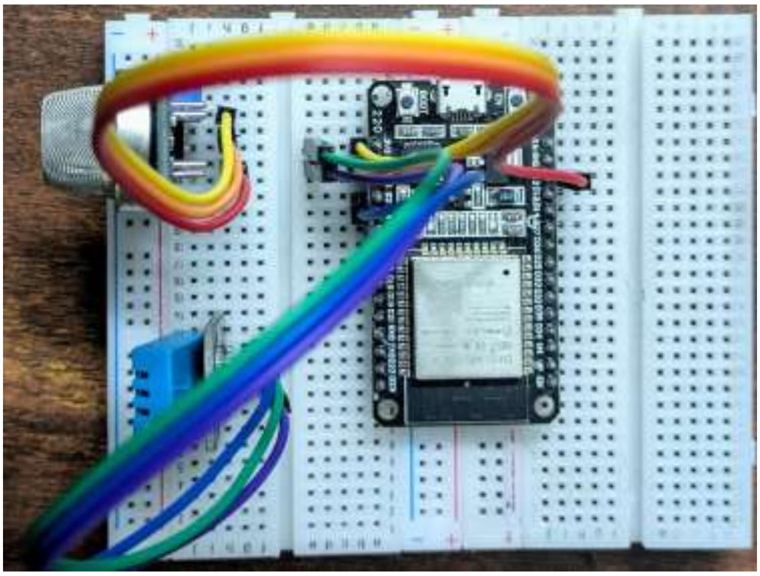

Figure 2: Prototype for Non-invasive Glucose sensor

Results:

1.4 Early Prediction

i. The model achieved accuracy of 0.98, precision of 0.99, recall of 0.98, and F1 score of 0.98.

ii. Loss and accuracy graphs (Fig. 3 and Fig. 4) indicate model performance with separate "Train" and "Test" datasets. The training accuracy was perfect at 1.0, and the testing accuracy reached 0.98.

iii. The proposed model outperforms existing algorithms, such as XGBoost, which showed fluctuating accuracy from 74% to 98.9% and Jaiswal et al. with 85% accuracy.

iv. With better statistical results (Accuracy: 0.98, Precision: 0.99, F1 score: 0.98, Recall: 0.98, RMSD: 0.02), the model is shown to be highly effective for early T2DM prediction, offering significant advantages for real-world applications.

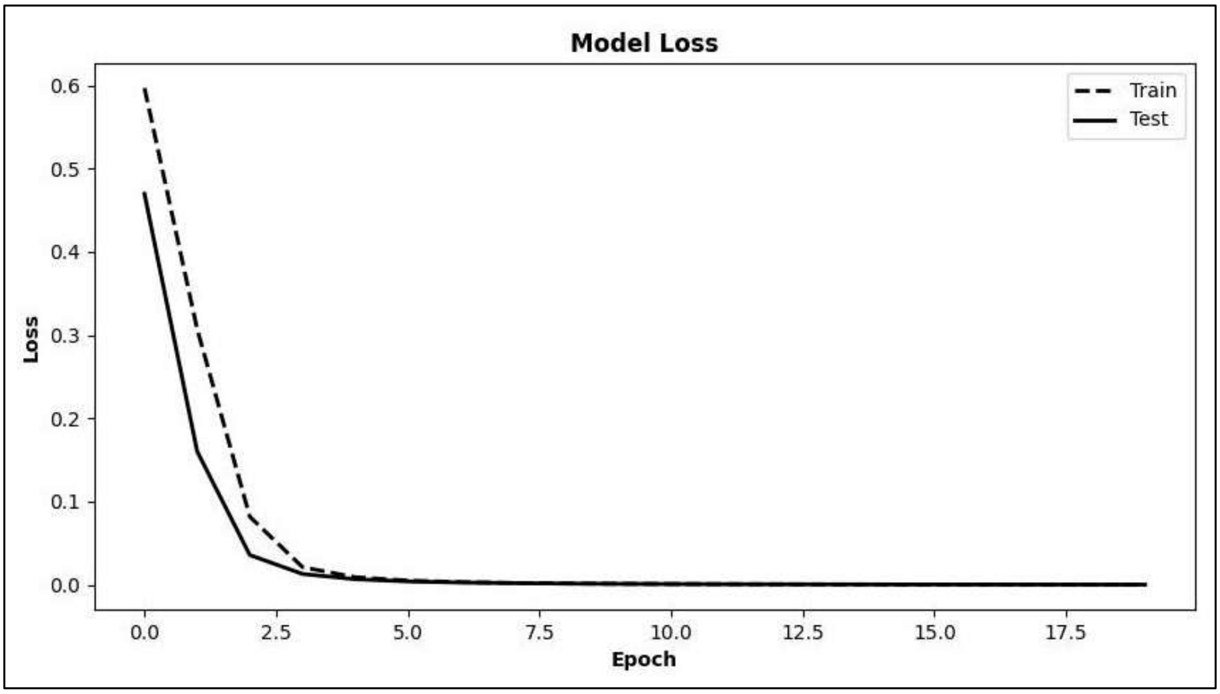

Figure 3: Error Analysis of LMST Model Outputs

.

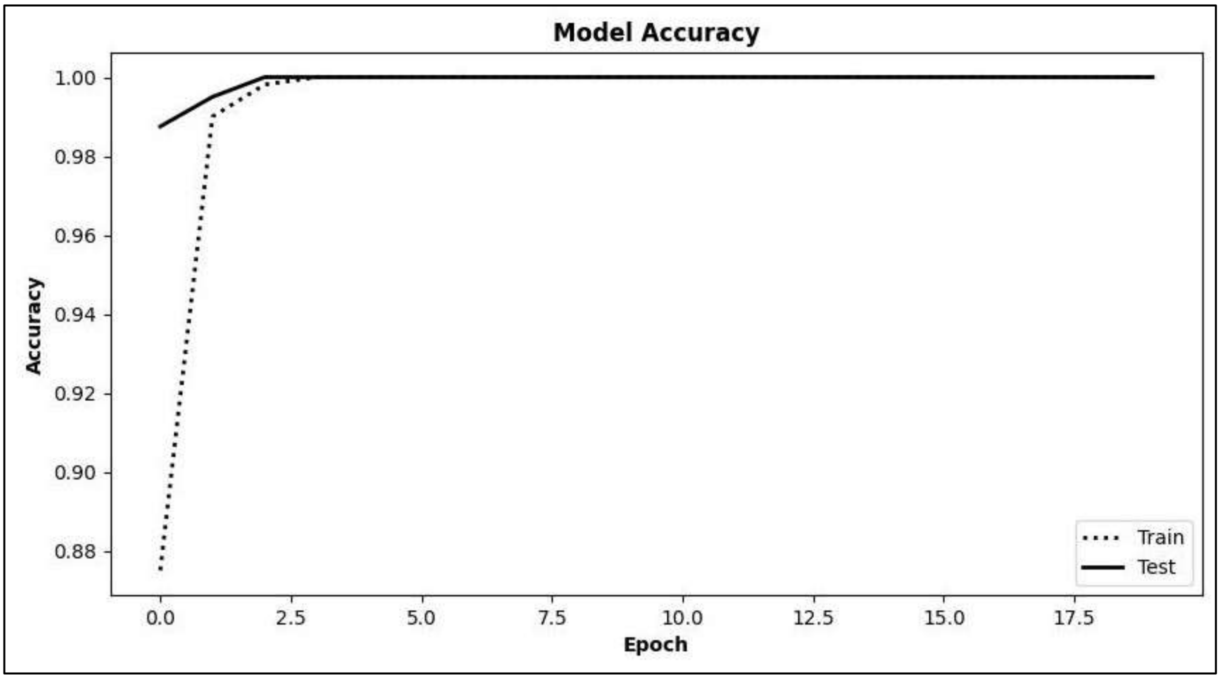

Figure 4: Accuracy Evaluation of LMST Model Outputs

Sensor-Based Monitoring

i. The IoT-based system, utilizing MQ135 and DHT11 sensors with the ESP32 microcontroller, successfully collected breath samples for glucose estimation.

ii. Multiple breath samples were gathered to ensure accurate readings, with stability checks applied to validate sensor consistency.

iii. The system provided reliable glucose estimations based on the collected breath data, demonstrating the potential for non-invasive monitoring.

iv. However, at this stage, no health recommendations or classifications were generated, as the focus was on gathering and validating sensor data for future integration with machine learning models.

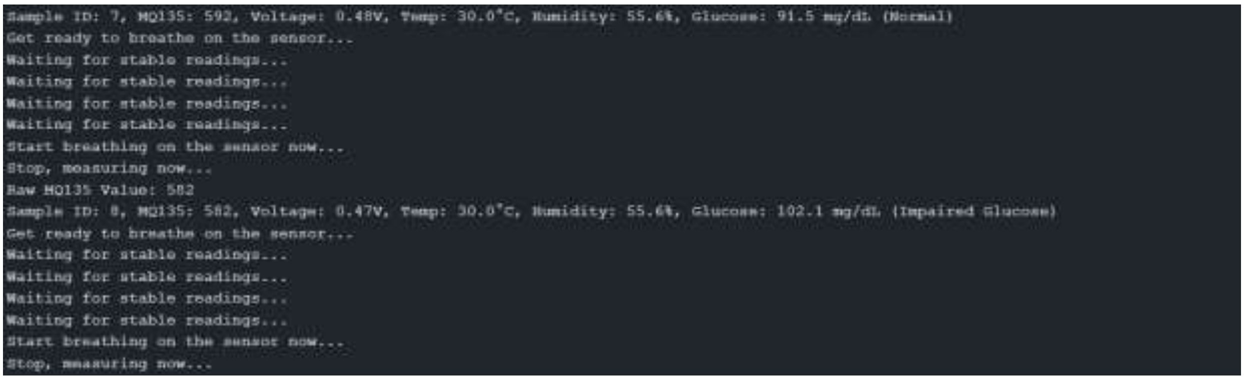

Figure 5: Sensor Results

Outcomes

1.5 A Paper has been published in the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT). The study leverages Long Short-Term Memory (LSTM) networks, a type of deep learning model, to predict Type-II Diabetes Mellitus (T2DM) in young adults with high accuracy.

Figure 6: Paper Published

1.6 The design patent granted for our innovative LSTM-based diabetes risk prediction system. The patent protects the visual and functional structure depicted, which systematically processes input features like glucose, BMI, and age through an LSTM model to classify diabetes risk levels.

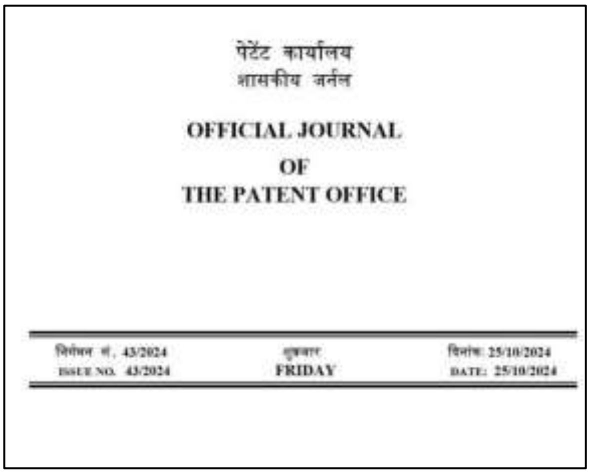

Figure 7: Patent Granted

1.7 The code developed for implementing an LSTM-based machine learning model has been copyrighted.

Extracts from the Register of Copyrights

Copyright Office, Government Of India

L-158396/2024

Figure 8: Copyright

Conclusion and Future Scope:

The increasing prevalence of Type 2 Diabetes Mellitus (T2DM) among young adults necessitates innovative and accessible diagnostic solutions. Our study demonstrates the effectiveness of machine learning techniques, particularly deep learning models, in accurately predicting the early onset of T2DM. The integration of IoT sensors for non-invasive breath analysis provides a novel approach to health monitoring, with the potential to reduce dependence on traditional, invasive diagnostic methods.

The proposed system has shown high accuracy in early prediction and consistent data collection through sensor-based monitoring. Though the initial prototype focuses on data acquisition without generating health recommendations, it lays the foundation for future development.

Future Scope:

I. Integrating the sensor system with the trained ML model to deliver real-time risk analysis and personalized recommendations.

II. Enhancing the device for continuous monitoring and mobile health integration (e.g., app-based interface).

III. Expanding the dataset to include diverse population samples for improved model generalization.

IV. Potential collaboration with healthcare providers for clinical validation and deployment in real-world scenarios.