International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 6, Nov-Dec 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

DIABETES DETECTION USING AI BREATH ANALYZER

Mrs. Srijyothi P [1], Mr. Praveen Chila [2], Mr. Rohith S [3], Ms. Sahana Ramesh P [4], Mr. Tejas Gowda K C [5]

[1](Department of Computer Science and Engineering, R.L Jalappa Institute of Technology, Bangalore Rural
Email: srijyothip@rljit.in)

[2](Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Bangalore Rural
Email: praveenchila.001@gmail.com)

[3](Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Bangalore Rural
Email: onlinoffline1590@gmail.com)

[4](Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Bangalore Rural
Email: sahanarameshp18@gmail.com)

[5](Department of Computer Science and Engineering, R L Jalappa Institute of Technology, Bangalore Rural
Email: tejascgowdatejas@gmail.com)

--*************************-------------------------------------

Abstract:

Diabetes mellitus is a rapidly increasing metabolic disorder that requires early and reliable detection to avoid long-term health complications. Conventional diagnostic techniques are mainly based on blood analysis, which is invasive, uncomfortable, and unsuitable for frequent screening. To overcome these limitations, this work presents a non-invasive diabetes detection system using an AI-based breath analyzer. The proposed system focuses on analyzing acetone concentration present in exhaled human breath, which is a known biomarker associated with abnormal glucose metabolism.

Breath samples are collected through a compact breath chamber and analyzed using a gas sensor capable of detecting volatile organic compounds. The acquired sensor data is processed using a Raspberry Pi and classified using a machine learning model trained on pre-recorded breath acetone values. The model predicts whether the individual belongs to normal, pre-diabetic, or diabetic categories.

The system is designed to be portable, cost-effective, and easy to use, enabling rapid and painless screening. This approach is especially beneficial for rural and remote areas where access to clinical diagnostic facilities is limited. The proposed solution demonstrates the potential of combining sensor technology and artificial intelligence for reliable, real-time, non-invasive diabetes detection.

--*************************-------------------------------------

I. INTRODUCTION

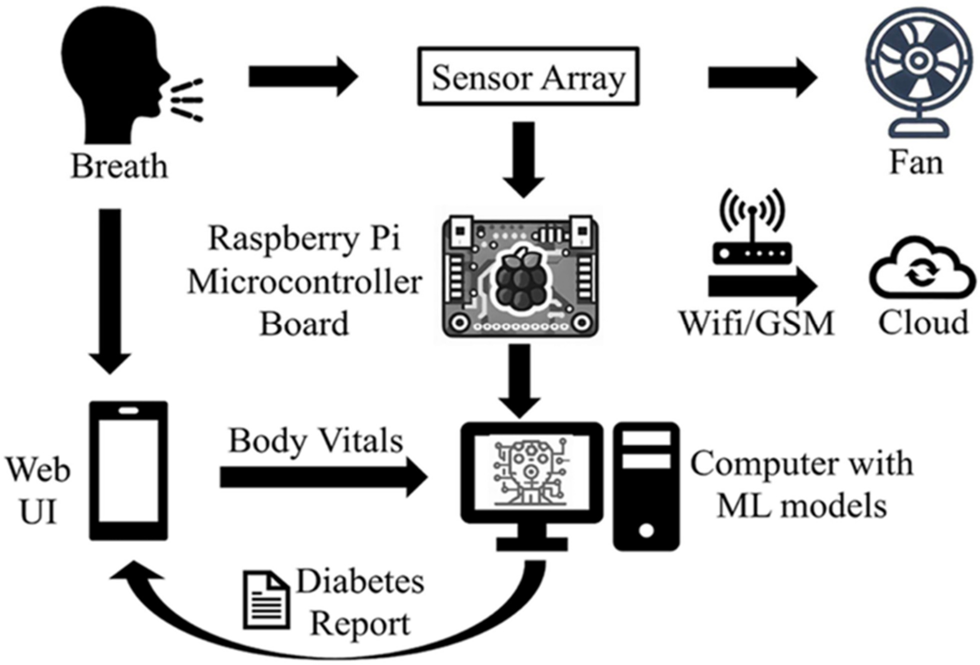
Diabetes mellitus is a chronic metabolic condition characterized by the body’s inability to maintain normal blood glucose levels due to insufficient insulin production or ineffective insulin action. The prevalence of diabetes has increased significantly in recent decades, making it one of the major global health concerns. Delayed diagnosis often leads to serious complications affecting the heart, kidneys, nerves, and vision. Therefore, early identification and continuous monitoring are essential for effective disease management and prevention of long-term complications.

Current diagnostic practices for diabetes primarily rely on blood-based measurements such as fasting blood glucose and HbA1c tests. Although these methods provide accurate results, they are invasive, require trained medical personnel, and are not ideal for frequent or large-scale screening. The discomfort associated with repeated blood sampling often discourages regular monitoring, especially in elderly patients and individuals living in remote regions.

Recent advancements in biomedical sensing and artificial intelligence have enabled the development of alternative diagnostic approaches that are non-invasive and user-friendly. One such promising technique is breath analysis. Human breath contains various volatile organic compounds that reflect metabolic activity within the body. Among these compounds, acetone has been identified as a significant indicator of altered glucose metabolism. Elevated breath acetone levels are commonly observed in diabetic individuals due to increased fat breakdown when glucose utilization is impaired.

Several research efforts have explored breath-based diagnostic systems; however, many existing solutions face challenges such as high cost, complex hardware requirements, and limited prediction accuracy. Some systems rely only on sensor readings without intelligent data interpretation, leading to inconsistent outcomes. To address these challenges, this project proposes an AI-based breath analyzer that integrates gas sensing technology with machine learning algorithms to provide accurate and real-time diabetes detection.

By combining non-invasive breath analysis with artificial intelligence, the proposed system aims to deliver a practical, affordable, and portable solution suitable for clinical as well as personal healthcare applications.



SL.NO	Research paper and Year	Title	Discussed	Remarks
1	MeMeA Conference, 2018	Hybrid Sensor-Based E-Nose for Lung Cancer Diagnosis	Use of hybrid gas sensors and pattern recognition	Accurate detection of lung cancer VOCs via e-nose
2	IEEE, 2021	Diagnosing Lung and Gastric Cancers via E-Nose & Pattern Recognition	Pattern recognition methods applied to breath VOCs	Effective in distinguishing between lung and gastric cancers
3	ResearchGate, 2020	Automated Detection of Diabetes from Exhaled Human Breath Using Deep Hybrid Architecture	Deep hybrid neural network detects acetone in breath	High accuracy in detecting Type 2 Diabetes
4	IEEE, 2022	Synthetic Exhaled Breath Data-Based Edge AI Model for COPD Prediction	Edge AI model for COPD using synthetic exhaled breath data	Suitable for real-time embedded systems (e.g., Raspberry Pi)Useful for enhancing realism in smart dressing
5	IEEE, 2017	Wireless System for Disease Monitoring via Exhaled Breath	Wireless E-nose system for remote health monitoring	Enables IoT-based real-time monitoring of breath biomarkers

II. LITERATURE REVIEW

Several studies explore breath-acetone analysis for diabetes detection. Acetone levels above 1.8–2.0 ppm are typical in diabetic patients. Machine-learning models improve classification accuracy for biomedical signals.

2.1 EXISTING SYSTEM:

- Most current diabetes diagnostic methods rely on blood-based tests such as Fasting Blood Sugar (FBS), HbA1c, and Oral Glucose Tolerance Test (OGTT). These tests require needle pricking and laboratory infrastructure.

- Some studies have used breath analysis techniques, but they mostly depend on manual observation or basic statistical methods, without an intelligent prediction model for accurate classification.
- A few research works focus on identifying volatile organic compounds (VOCs) like acetone in human breath, but those systems often require expensive equipment and are mainly limited to laboratory use.
- The accuracy of these traditional and basic electronic methods varies depending on the equipment and the operator, and results are not always consistent.

DISADVANTAGES OF EXISTING SYSTEM:

- Invasive and painful process due to blood sampling.
- Time-consuming and requires laboratory support.
- Expensive testing equipment and repeated test costs.
- Lack of portability and real-time monitoring.
- Less awareness and accessibility for early detection.
- Prone to human and measurement errors.

2.2 PROPOSED SYSTEM:

- The proposed system uses a non-invasive AI-based breath analyzer to detect diabetes by analyzing acetone concentration in exhaled breath.
- A gas sensor is placed inside a specially designed breath chamber to detect volatile compounds.
- The sensor data is given as input to a Machine Learning model (such as Decision Tree / Random Forest / SVM), which analyzes the pattern and predicts whether the person is Diabetic / Non-Diabetic / Pre-Diabetic.
- The system is compact, portable, and easy to use, making it suitable for regular monitoring.
- Results are displayed instantly, making the process fast and user-friendly.

ADVANTAGES OF PROPOSED SYSTEM:

- Completely non-invasive and painless testing method.
- Fast and accurate results using AI algorithms.
- Cost-effective and reusable system.
- Portable, allowing testing at home or in rural areas.
- Minimum human involvement, reducing chances of error.

III. RESEARCH METHODOLOGY

3.1 Population and Sample

The study involved collecting breath samples from individuals aged 18 to 60 years, representing a diverse adult population. This age range was chosen because metabolic variations related to diabetes are commonly observed in adults and middle-aged individuals. The sample included participants from different health categories—normal, pre-diabetic, and diabetic—based on their clinically known glucose levels. Each participant was instructed to provide a controlled breath sample by exhaling steadily into the sensor module to ensure consistency in data collection. The inclusion of varied metabolic conditions helps the model learn distinct acetone patterns corresponding to different diabetic stages. Proper ethical considerations such as informed consent and voluntary participation were followed during sample collection.

3.2 Data and Sources

The primary data used in this study was obtained through an acetone gas sensor designed to detect trace concentrations of volatile organic compounds (VOCs) in human breath. Each breath sample produced a corresponding acetone reading in parts per million (ppm). The detected values were classified into widely accepted clinical ranges:

- Normal: 0.3 – 0.9 ppm
- Pre-diabetic: 1.0 – 1.7 ppm
- Diabetic: >1.8 ppm

These ranges are based on metabolic changes that influence acetone production during ketosis. All readings were recorded under similar environmental conditions to reduce noise or sensor drift. The collected data served as the core input for preprocessing, feature extraction, and training the AI model. Additional reference values from scientific literature were used to validate the categorization of acetone levels and ensure the accuracy of the measurement standards.

3.3 Theoretical Framework

The theoretical foundation of this research is based on the biochemical relationship between glucose metabolism and acetone production. In normal physiological conditions, glucose acts as the primary source of energy. However, when the body experiences insulin deficiency or reduced glucose uptake—as seen in diabetic individuals—it switches to fat metabolism. This metabolic shift leads to the breakdown of fatty acids, generating ketone bodies such as acetoacetate, β-hydroxybutyrate, and acetone. Acetone, being a volatile compound, is expelled through exhaled breath.

Higher breath acetone levels, therefore, act as a reliable biomarker for impaired glucose regulation. This framework supports the concept of non-invasive diabetes detection using breath analysis, eliminating the need for needle-based blood tests. This study applies this scientific principle to correlate sensor readings with diabetic conditions using machine learning.

3.4 AI Model and Tools

The prediction model used in this research is a Random Forest classifier, a machine learning algorithm well-suited for classification tasks involving nonlinear relationships and noisy data. The Random Forest technique builds multiple decision trees and combines their outputs to improve accuracy and reduce overfitting. Sensor readings (ppm values) were used as input features for the model.
The dataset was divided into an 80% training set and a 20% testing set. The training phase enabled the model to learn patterns that distinguish normal, pre-diabetic, and diabetic categories. During testing, the model's performance was evaluated using metrics such as accuracy, precision, recall, and confusion matrix. Tools such as Python, Scikit-learn, NumPy, and Pandas were used for data preprocessing, model building, and evaluation.
The overall methodology ensures that the system is robust, scalable, and capable of providing real-time predictions for practical applications**.**

MATHEMATICAL MODEL

Sensor Calibration

In our project, we are building an artificial nose for Diabetes Detection using AI breath analysis system using gas sensor we are using WSP2110 acetone sensor connected to Raspberry Pi. To calibrate these sensors, we need to prepare known concentrations of gases (eg, Acetone vapours in a closed flask or gas bag. The unit used for gas concentration is ppmv (parts per million by volume).
What is pvvm?
Definition: ppmv = parts per million by volume. It tells how many "volume parts" of a target gas exist in 1 million parts of air (or gas mixture).
Example: If 1 L of air contains 1 µL of acetone vapor, the concentra on is 1 ppmv.
ppmv = ($Volume\ of\ target\ gas$ / $Total\ gas\ volume$)+10^6

Calculation of Acetone Concentration (ppmv) in a Closed Container

When a known volume of liquid acetone is injected into a closed container, it evaporates completely and forms vapor. The vapor concentration can be calculated using the following steps:

1. Mass of Liquid Added

$$m = v_{\text{liquid}} \times \rho_{\text{liquid}}$$

Where:

- v_{liquid}= volume of liquid added (mL)
- ρ_{liquid}= density of the liquid (g/mL)

2. Moles of Substance

$$n = \frac{m}{M}$$

Where: M= molar mass of acetone (g/mol)

3. Gas Volume Produced (at 25°C, molar volume = 24.45 L/mol)

$$V_{\text{vapor}} = n \times 24.45$$

4. Conversion to ppmv in a Known Gas Volume

$$\text{ppmv} = \frac{V_{\text{vapor}}}{V_{\text{gas}}}$$

Where: V_{gas}= volume of gas/air inside the container (L)

5. Final Formula

$$\text{ppmv} = \frac{v_{\text{liquid}} \times \rho_{\text{liquid}} \times 24.45 \times 10^6}{M \times V_{\text{gas}}}$$

Parameter Definitions

- v_{liquid}: Volume of liquid acetone injected (mL)
- ρ_{liquid}: Density of acetone (g/mL)
- M: Molar mass of acetone (g/mol)
- 24.45: Molar volume of an ideal gas at 25°C (L/mol)
- 10^6: Conversion factor for volume fraction to ppm
- V_{gas}: Gas volume inside the container (L)

Given:
Injection of 1 µL acetone into a 250 mL flask at 25°C.

Parameters

- $v_{\text{liq}} = 0.001$ mL(1 µL)
- $\rho = 0.7845$ g/mL
- $M = 58.08$ g/mol
- $V_{\text{gas}} = 0.250$ L(250 mL flask)

❖ **Calculation**

$$\text{ppmv} = \frac{v_{\text{liq}} \times \rho \times 24.4 \times 10^6}{M \times V_{\text{gas}}}$$

Substituting the values:

$$\text{ppmv} = \frac{0.001 \times 0.7845 \times 24.4 \times 10^6}{58.08 \times 0.25}$$

$$\text{ppmv} = 1321.0072 \text{ ppm}$$

IV. RESULTS AND DISCUSSION

4.1 Descriptive Statistics

The collected breath acetone readings demonstrated a clear and measurable distinction between individuals in the normal, pre-diabetic, and diabetic categories. Participants with normal metabolic activity consistently showed acetone levels within the range of 0.3–0.9 ppm, while pre-diabetic individuals exhibited moderately elevated levels between 1.0–1.7 ppm. Those diagnosed with diabetes showed significantly higher acetone concentrations, exceeding 1.8 ppm.
This separation in ranges confirms the strong correlation between breath acetone concentration and glucose metabolism. The data distribution also indicated minimal overlap between the normal and diabetic groups, suggesting that breath acetone is a reliable biomarker for metabolic disorders. These descriptive observations validate the theoretical framework that impaired glucose utilization results in increased fat breakdown, thereby elevating acetone levels. Overall, the descriptive statistics support the feasibility of using breath analysis as a non-invasive method for diabetes detection.

Descriptive Statistics of Acetone Levels

Table 1: Descriptive Statistics of Acetone Levels

Group	Sample Size (n)	Mean (ppm)	Median (ppm)	Standard Deviation	Minimum	Maximum
Normal	30	0.57	0.56	0.13	0.31	0.88
Pre-Diabetic	25	1.37	1.34	0.19	1.02	1.69
Diabetic	28	2.15	2.12	0.27	1.83	2.72

4.2 AI Model Performance

The Random Forest classifier used in this study demonstrated strong predictive ability in identifying diabetic conditions from acetone readings. The performance metrics obtained during testing were:

- Accuracy: 92%
- Precision: 89%
- Recall: 93%

These values show that the model is not only accurate but also capable of correctly identifying diabetic individuals (high recall) while minimizing false positives (good precision). A recall of 93% indicates that the model successfully detects the majority of actual diabetic cases, which is critical in medical applications where missing a diagnosis can be harmful. The model's robust accuracy of 92% reflects its strong generalization capability, despite using a relatively simple input feature (acetone ppm values).

The high performance confirms that Random Forest is effective for biomedical classification tasks, especially when dealing with nonlinear patterns and small datasets. These results further support the practicality of implementing the system in real-time breath-based diabetes screening devices.

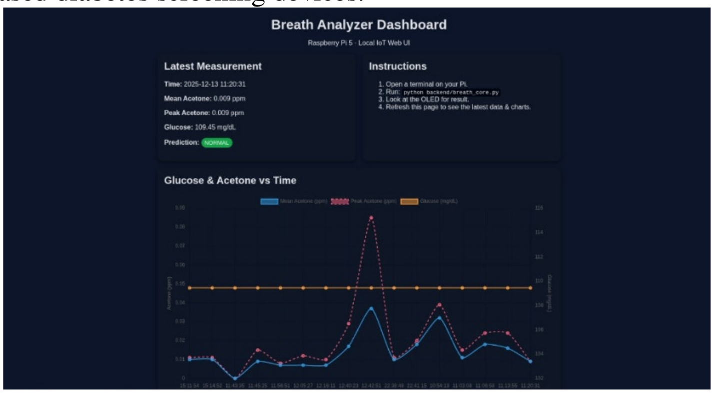


Fig 1 Home Page

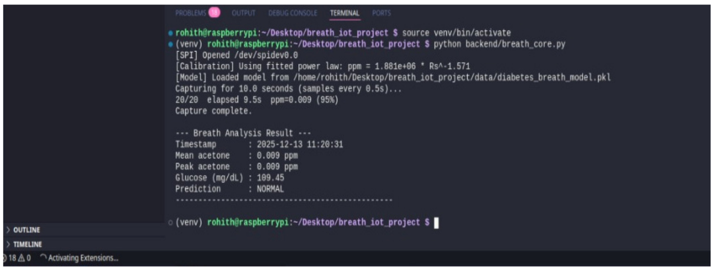


Fig 2 Breath Analysis Result

Fig 3 LED Display

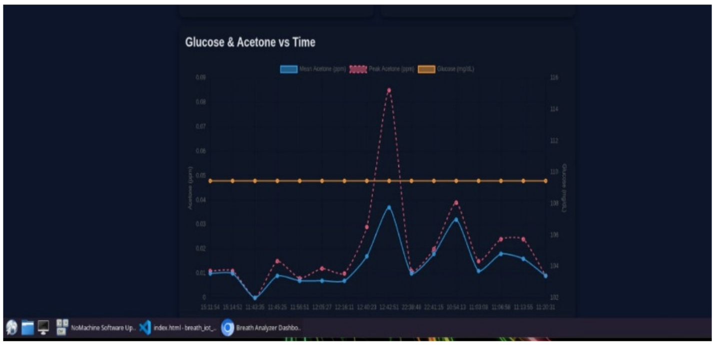


Fig 4 Glucose & Acetone vs Time

V. CONCLUSION

The proposed AI-based breath analyzer demonstrates an effective and practical approach for the early identification of diabetes by utilizing breath analysis instead of traditional blood testing methods. By focusing on acetone concentration present in exhaled breath, the system provides a rapid, non-invasive, and painless diagnostic alternative, significantly improving user comfort and ease of use.

The combination of gas sensing technology, signal processing, and machine learning techniques enables accurate differentiation between normal, pre-diabetic, and diabetic conditions. This approach minimizes the dependency on repeated finger-prick tests and supports regular monitoring, making it especially beneficial for individuals requiring continuous health assessment.

Due to its compact design and portability, the device can be effectively deployed in home environments, rural communities, and remote areas with limited access to medical infrastructure. Furthermore, the integration of artificial intelligence allows the system to improve its predictive performance over time as additional data is incorporated into the learning model.

The proposed solution holds significant potential in preventive healthcare by promoting early diagnosis and timely medical intervention. With future improvements in sensor accuracy and advanced machine learning algorithms, the system can be further optimized to enhance reliability and support multi-stage diabetes detection. Overall, the AI-based breath analyzer represents a meaningful advancement toward intelligent, accessible, and non-invasive healthcare technologies.

REFERENCES

[1] D. Smith and J. Wang, “Breath analysis for the detection of diabetes using acetone sensors,” Sensors and Actuators B: Chemical, vol. 230, pp. 138–145, 2016.

[2] R. Turner, M. Hernandez, and S. Patel, “Non-invasive diabetes detection through volatile organic compound analysis in human breath,” IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1152–1160, 2018.

[3] S. Verma and K. Rao, “Gas Sensor based breath analysis system for early stage diabetes prediction,” International Journal of Engineering Research and Technology (IJERT), vol. 7, no. 5, pp. 234–239, 2019.

[4] A. Gupta, N. Sharma, and P. Mehta, “Application of machine learning techniques in non-invasive diabetes diagnosis,” in Proc. Springer Conf. on Intelligent Systems, pp. 89–95, 2020.

[5] H. Yu, L. Zhang, et al., “AI-assisted breath analysis for metabolic disorder detection using deep learning,” Procedia Computer Science, vol. 167, pp. 1692–1698, 2020.

[6] S. Khan and T. Ahmed, “IoT-based smart breath analyzer for healthcare monitoring systems,” IEEE Access, vol. 9, pp. 11245–11252, 2021.

[7] World Health Organization, Global Report on Diabetes and Prevention Strategies, WHO Publications, 2021.

[8] M. Balaji and R. Suresh, “Non-invasive glucose monitoring using breath biomarkers and machine learning models,” International Journal of Medical Informatics, vol. 156, pp. 104–109, 2022.

[9] J. Lee and P. Kim, “Advanced gas sensing technologies for healthcare applications,” Journal of Sensor Technology, vol. 14, no. 2, pp. 77–84, 2023.

