I SSN (Print):2663-2381
ISSN(Online):2663-4007
| Regular Article | Open Acess
Aravandi Publications

Breathalyzer 2.0: Non-Invasive Diabetes Detection Using Machine Learning

Mahalakshmi K*[1], Jelcy Chiristila P[2], Dr. Shanmugavalli M[3], Preethy Janet P[4]

[1,]Student, Department of Instrumentation and Control Engineering, Saranathan College of Engineering, Tamil Nadu, India.
[2]Student, Department of Instrumentation and Control Engineering, Saranathan College of Engineering, Tamil Nadu, India.
[3]Professor, Department of Instrumentation and Control Engineering, Saranathan College of Engineering, Tamil Nadu, India.
[4]Professor, Department of Computer Science and Engineering, Saranathan College of Engineering, Tamil Nadu, India.

Corresponding author(s): DoI: https://doi.org/10.5281/zenodo.18308653
Mahalakshmi K, Student, Department of Instrumentation and Control Engineering, Saranathan College of Engineering, Tamil Nadu, India.
Email: kkmsrimaha@gmail.com

Citation:

Mahalakshmi K, Jelcy Chiristila P, Dr. Shanmugavalli M, Preethy Janet P (2026). Breathalyzer 2.0: Non-Invasive Diabetes detection using Machine Learning. International Journal of Multidisciplinary Research Transactions, 8(1), 1–13. DoI: https://doi.org/10.5281/zenodo.18308653

Accepted: 18 January 2026 Available online: 20 January 2026

Abstract

The Diabetic Patient count is drastically increased nowadays, Diabetes is a long term condition in which the person's body cannot breakdown the blood sugar adequately due to shortage in the production of insulin. Diabetes should be monitored properly to reduce critical emergencies such as Hyperglycemia and Hypoglycemia. Previous method used to detect diabetes by pricking fingers and by testing it with glucometer which is a invasive method for continuous monitoring, frequent pricking will become painful process. To avoid this, non-invasive method is used. Non Invasive Diabetes detection based on the variation of Acetone level from the exhale breath of the patient. Volatile organic compound(VOCs) such as acetone, ethanol, toluene, Benzene, formaldehyde causes variation in blood glucose level and insulin level. Acetone from the exhale breath provides a direct changes in blood glucose and insulin levels. From that the blood glucose variation can be measured directly based on the change in acetone level .The research discuss on non-invasive techniques to measure the blood glucose level from the exhale breath. Machine learning model is trained using support vector regression(SVR) with a small range of dataset to provide a higher accuracy and to reduce error in the outcome.

Keywords: Acetone sensor, Exhale breath analysis, Non-invasive method, ML model Training(SVR), Volatile organic compound(VOCs) ACETONE.

1. Introduction

Diabetes is common in many people's nowadays caused due to the improper insulin secretion in human body. The glucose level is monitored using a Non-invasive method to detect and measure the blood glucose level from the acetone concentration in ppm from the exhaled breath of human. Exhale breath contains VOCs such as ethanol, methanol, benzene, formaldehyde, acetone, Toluene etc… the major compound used is acetone, to measure the variation in blood glucose level and to detect the diabetes without the blood need and also non-invasive method is suitable for continuous monitoring.

1.1. Literature Survey

Diabetes is long term condition in which person body cannot breakdown the blood glucose to glycogen which is measured using non-invasive measurement technique for continuous monitoring without the requirement of blood samples, and also highlights on traditional invasive techniques and emphasize Smart healthcare framework is built on Internet of Medical things(IoMT) and a Healthcare Cyber Physical System(H-CPS) also it identifies major challenges and issues related to precise, calibrated and predict the glucose monitoring system [1]. Volatile Organic Compound such as Acetone is used to diagnose the diabetes through non-invasive method from the exhaled breath of the patient to avoid frequently pricking the fingers of the patient for continuous monitoring [2].D-Shaped composite plasmonic sensors (DSCPS) Sensor response is characterized by a full vectorial finite element method(FEM)[3]. Sensor field mainly focuses on W03, ZnO ,SnO2, Fe2O3, TiO2, etc..[4]. Diabetes is a metabolic disorder which cannot regulate the blood glucose level properly this occurs due to improper insulin secretion, which is produced by the pancreas, Data extracted from sensor uses Principle component analysis(PCA) and Linear discriminant analysis(LDA)[5]. As the insulin should be properly produced to convert the glucose to glycogen (Stored in Liver and Muscles). If the body cannot produce insulin properly then it is consider as Type-1 Diabetes and if the body cannot use the insulin properly for the glycogen conversion then it is considered as Type-2 Diabetes

(Insulin Resistance). Due to high glucose level in blood causes damage in blood vessels, nerves, and other organs[6]. A Multi wave-length photoplethysmography (PPG) used to improve the non-invasive blood glucose measurement by calculating amplitude ratios (R/G and IR/G) using green light which is a baseline for calibrating blood flow measurement in the total area and the results are demonstrated using Simple linear regression model to attain 90 % accuracy in Clarke plot analysis highlighting clinical applicability of approach also AI techniques could improve the predictivity and accuracy of the model[7].

In this paper, A non invasive blood glucose measurement through exhale breath is discussed in section 2. Section 3 explains about the simulation result of the non invasive glucometer and Section 4 describes about the ML model training and displaying blood glucose level. Section 5 visualizes the hardware connection and final predicted glucose level. Section 6, discuss about the glucose level prediction as an output and the paper is finally concluded in section 7.

2. Measurement Techniques Used In Glucometer

This experiment mainly focuses on developing a real time blood glucose prediction from the acetone concentration (in ppm) in exhale breath. This technique involves combination of hardware and software integrating, which involves sensor based data acquisition, data logging, and machine-learning based prediction. This experiment ensures accurate data collection and reliable glucose estimation. The acetone concentration from the exhaled breath is detected using MQ-138 which is sensitive to volatile organic compounds (VOCs) including acetone.

Here, acetone is used as a major source to detect blood glucose level. This sensor works on the principle of change in resistance when exposed to acetone in the exhaled breath. The Analog signal from the sensor is given to Arduino UNO microcontroller (ATMEGA328P) for continuous voltage output. The Arduino board is programmed to convert the Analog signal into digital form and which is stored in csv file for direct comparison between the trained ML model and input data to produce the predicted output. The data transmission ensures accurate monitoring of the data. A python program is used to receive the sensor reading from the Arduino and which is stored in the csv file automatically.

Each patient record include a timestamp and the corresponding voltage or sensor output. This creates a live dataset that indicates the acetone concentration. The data collected from the is used as a input and which is compared with the pre-trained ML model that correlates the acetone concentration with the blood glucose level. The Pre-trained ML model is based on the Support vector regression(SVR). During the live performance the python program continuously read the latest input from the csv file, converts sensor voltage to acetone concentration using calibration factors. Then, the data is sent to the ml model to predict the glucose level, also the data can be displayed on the monitor and stored in the csv file for future use. This setup is a low-cost, non-invasive app roach for glucose estimation. Fig 1: explains about the working flow of non-invasive glucometer and predicted blood glucose range. Table 1: discuss about the different glucose range of diabetes patient before fasting and after fasting.

3. Acetone And Blood Glucose Correlation

Table.1. Blood Glucose Range of Diabetic Patient

Time	Condition	Measured Acetone (ppm)	Estimated Glucose(mg/dL)
Morning(after breakfast)	Before fasting	0.6	177
After 6 hours	Mid-fasting	1.8	170
After 12 hours	Full-day fast	3.5	162
After 24 hours	Extended fast	5.2	153

4. Block Diagram Of Breath Glucometer

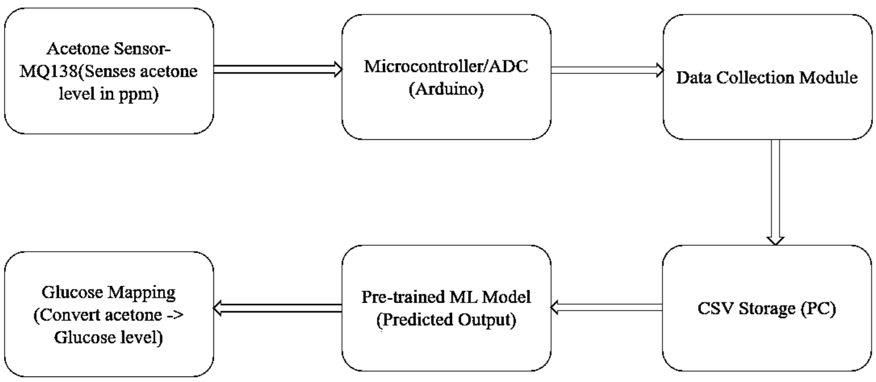

Figure.1. Block Diagram Of Breath Glucometer

5. Simulation Of A Non Invasive Glucometer

For this, Simulation output of the Breathalyzer which predicts the blood glucose level through non-invasive method which can be done by collecting the acetone level from the exhaled breath.

Which can be performed using the combination of Arduino IDE, and python based machine learning tools which involves packages such as NumPy, Pandas etc… which shows the simulated environment before the hardware implementation. Here, Cirkit designer software is used to interface MQ-138 acetone sensor with Arduino UNO where the embedded c program is Dumped to collect data from sensor. Which is then saved as a csv file format, then the pretrained model and collected data are compared and the predicted output is displayed on the screen. Major components required are Acetone MQ-138 sensor Arduino UNO and jumper wires for connection and pretrained ML model. In fig 2: the simulation image gave an idea of the connection for the acetone (MQ138) sensor and data collection through Arduino which is connected to pre-trained ML model.

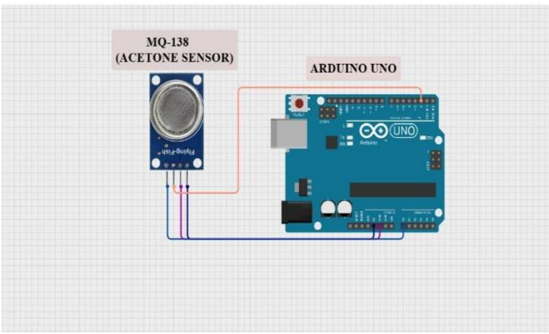

Figure.2. Simulation diagram of Acetone device

6. Ml Model Evaluation, Methodology And Discussion Of Breath Glucometer

6.1. Overview

This Python program implements a fully optimized machine learning system for diabetes detection using breath analysis data or other physiological indicators. It focuses on achieving a balanced accuracy of 85–90% without overfitting. This model integrates multiple ML techniques such as support vector regression(SVR), random forest, gradient boosting algorithm, and logistic regression to provide both regression and classification for predicting the data.

6.2. Main Class: Breathalyzer Final Optimized

The core of the program is 'Breathalyzer final optimized' , which encapsulate major processes including data loading, feature engineering, training the model, Visualizing and produces clinical interpretations. Which ensures design model, and each function perform a clear tasks.

6.3. Feature Engineering

The 'load and engineer features()' where the diabetes data are loaded , and creates data such as glucose, BMI ratio, insulin-to-glucose ratio and diabetes risk values or ranges. Which handles missing values by replacing zeros or with median values and categorizes age, BMI, and glucose levels into groups.

6.4. Feature Preparation

The 'Prepare features()' selects the relevant features to model and define three predicted targets:

- Blood Glucose Level (using regression is given in image in fig 3).
- Insulin Level (using regression is given in image in fig 4).
- Diabetes prediction (using Classification is given in image in fig 5). These various stages separates feature input and target output to train model.

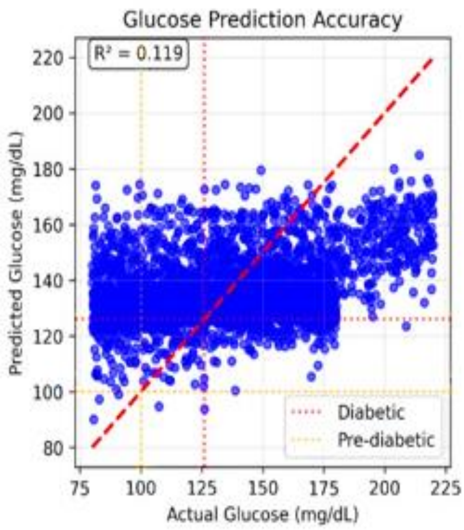

Figure.3. Glucose prediction

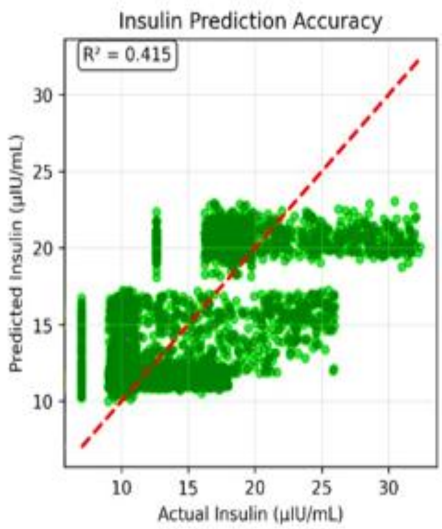

Figure.4. Insulin Prediction

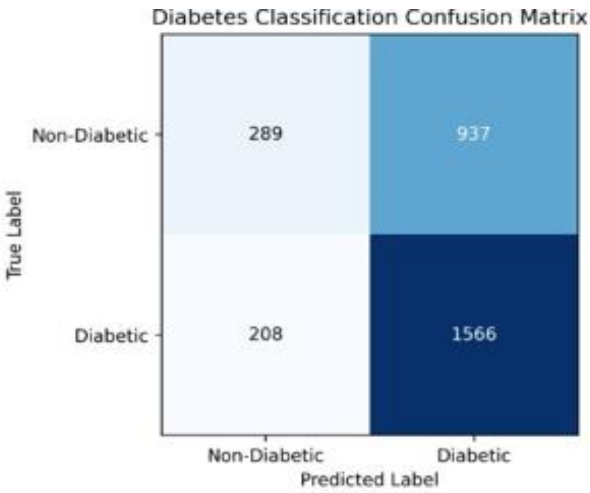

Figure.5. Diabetes classification

6.5. Optimized Model Training

The 'Train optimized model()' trains optimized model for each targets. It apply robust scaling to handle outlier and also uses feature selection with SelectKBest. For diabetes classifying ,it tune hyperparameters using GridSearchCV using multiple classifier such as SVC, Random forest, Gradient Boosting, Logistic Regression, combining them to achieve balanced and generalized performance.

For glucose and insulin prediction, SVR(support vector regression) and random forest regression for train both and selects the best performing model using cross validation Mean absolute error(MAE).

6.6. Evaluation Metrics

This program evaluates multiple metric using models:

- Accuracy, AUC, cross validation accuracy for classification.
- Mean absolute error(MAE), Root mean square(RMSE), and R^2 score for regression which also calculates an 'Overfitting Gap' to measure the generalization between training and test data.

6.7. Visualization

The `create final visualizations()` generates three scatter plots shows predicted vs actual values for:

- Diabetes classification(in fig 5 for visualization).
- Blood glucose prediction(in fig 7 for visualization).
- Insulin prediction(in fig 6 for visualization).

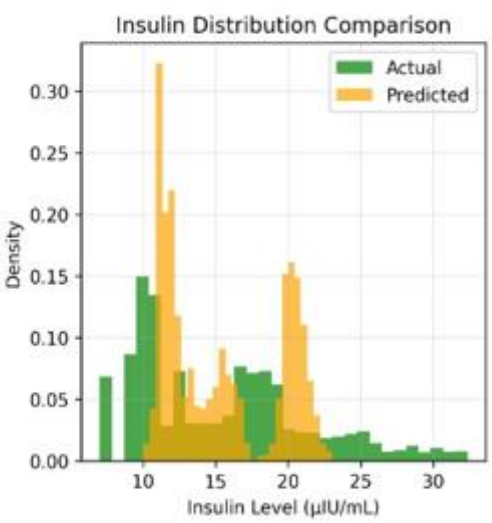

Figure.6. Insulin distribution

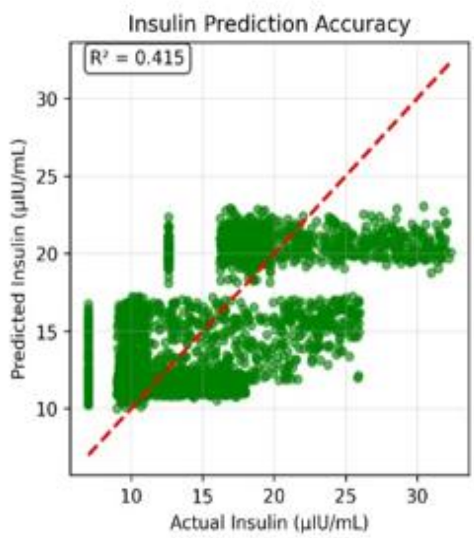

Figure.7. Insulin Accuracy

Each visualization includes performance factors such as accuracy, AUC, MAE, and R^2 displayed on the plot for interpretability.

6.8. Clinical Interpretation

The `clinical interpretation demo()` simulates clinical cases for normal, pre-diabetic, and diabetic patients. It provides a table showing glucose and insulin levels along with physiological meanings and clinical recommendations and their corresponding ranges. This connects machine learning predictions with real-world medical interpretation. Fig 8, Visualizes the clinical accuracy in graphical format for clear understanding.

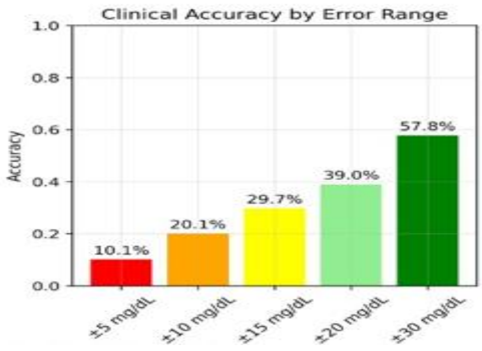

Figure.8. Clinical accuracy

6.9. Results and Model Saving

The 'Predict all samples to csv()' applies the trained model to all samples to the dataset, creates a CSV file as ('diabetes_with_prediction.csv') which include physiological interpretation for each sample data. The 'Save final model()' which stores the trained, scalers, selectors for future inference in the folder.

6.10. System Execution Flow

The `run final system()` function orchestrates all stages of the pipeline:

1. Load and engineer features
2. Prepare features
3. Train optimized models
4. Generate visualizations
5. Perform clinical interpretation
6. Predict all samples and creates CSV
7. Save final models

6.11. Summary

This system represents comprehensive, and medical interpretable AI pipelines to predict and monitor diabetes. Achieve strong generalization by balancing accuracy and to avoid overfitting, which combines multiple models.

The clinical interpretation features make it ready for integration into diagnostic or research applications.

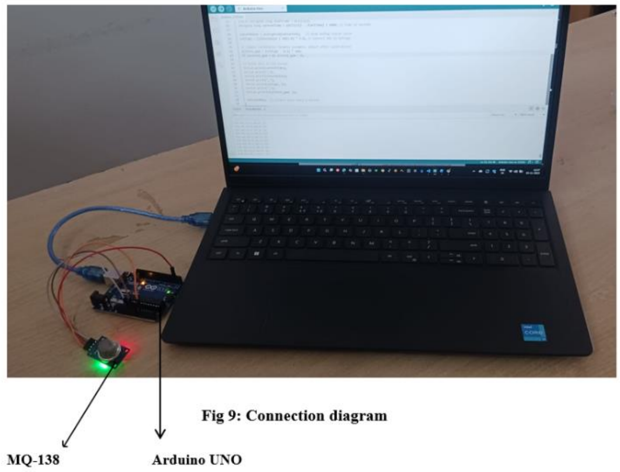

Figure.9. Connection of Sensor and Arduino

In fig 9,explains the connection of sensor and Arduino collects the acetone level from the patient and store the data in csv format which is compared with the pretrained ML model and displays the output in predicted csv file.

7. System Validation of Breath Glucometer

The result of the Breathalyzer clearly give the predicted glucose level effectively from the exhaled breath of the acetone concentration as a non invasive glucose levels. This project successfully integrates the real-time data storage, and machine learning prediction for providing a reliable glucose level. During the hardware experimentation MQ-138 acetone sensor produce different output voltage corresponding to different concentration of acetone collected from the exhaled breath sample. The readings are collected using Arduino UNO and converted into csv file communicated through serial communication which is compared with the pretrained ML model to predict the glucose level. Here, the observed data shows a clear inverse relationship between acetone and glucose levels, Physiological effect of this studies shows indicating diabetes patient with poor glucose regulation have higher acetone level in their exhale breath. This provide a feasible output using low-cost, non-invasive, using breath alternative approach can be applied to it , which offers a significant potential for diabetes health for continuous monitoring. In fig 10 and 11, shows the acetone level sensed from the sensor. Table 12 shows the predicted final output.

Time(s), Sensor value, Voltage(V), Acetone(ppm)

```
51,52.00,0.25,30.83
52,53.00,0.26,31.81
53,50.00,0.24,28.88
54,46.00,0.22,24.97
55,39.00,0.19,18.12
56,34.00,0.17,13.24
57,32.00,0.16,11.28
58,30.00,0.15,9.33
59,29.00,0.14,8.35
60,28.00,0.14,7.37
```


Figure.10. Data generated from the sensor (with low acetone level)

Time(s), Sensor value, Voltage(V), Acetone(ppm)

```
10,48.00,0.23,26.92
11,50.00,0.24,28.88
12,52.00,0.25,30.83
13,54.00,0.26,32.79
14,55.00,0.27,33.76
15,55.00,0.27,33.76
16,56.00,0.27,34.74
17,56.00,0.27,34.74
18,57.00,0.28,35.72
19,57.00,0.28,35.72
```

Figure.11. Data generated from the sensor (with high ppm acetone level)

Here, acetone level is show in larger value because of formulae used to convert in the program to obtain the sensor readings.

In the below figure 12, the predicted blood glucose output from acetone level is presented.

INITIALIZING BREATHALYZER FINAL OPTIMIZATION

Glucose	Insulin	Outcome	Predicted_Glucose_mgdL	Predicted_Insulin_uIUmL	Physiological_Meaning
148	125	1	149	125.0	Severe diabetes (DKA)
85	125	0	85	125.0	Severe diabetes (DKA)
183	125	1	182	125.0	Severe diabetes (DKA)
89	94	0	89	94.1	Severe diabetes (DKA)
137	168	1	141	168.1	Severe diabetes (DKA)

Figure.12. Final optimized (predicted) output of breath glucometer

8. Future Advancement In Breath Glucometer

This research focus on the new technique to determine the blood glucose level through breath based acetone detection.

- Nanomaterial based sensing materials which includes graphene functional metal oxides.
- Replacing the sensor MQ138 with TGS1820 for avoiding the environmental factors disturbance.
- The above sensing technique are more suitable to determine the level of acetone appropriately.
- Intelligent breath sampling strategies foe end tidal breath collection and adaptive humidity compensation techniques explored to minimize the environmental and physiological interference.

9. Conclusion

This Breathalyzer device for predicting blood glucose level from acetone concentration from exhale breath has been completely designed, simulated, and evaluated. The integration of sensor technology, embedded programing, and ML technique gives the non-invasive blood glucose monitoring in feasible and effective method to predict blood glucose level. MQ138 gas sensor used to detect acetone concentration in exhale breath, which is a indirect indication of blood glucose through metabolism change in human body. Continuous monitoring of sensor output using Arduino UNO and data collected and stored efficiently in CSV file to provide accurate blood glucose level. This provides a low-cost, portable non-invasive glucose monitoring with higher predictability also environmental affecting factors are pretrained to the model. Future improvements includes improvement in patient profile history, continuous diabetic monitoring data and smart healthcare system implementation.

Acknowledgement
The authors have no acknowledgements to declare.

Funding
This study has not received any funding from any institution/agency.

Conflict of Interest/Competing Interests
No conflict of interest.

Data Availability
The raw data supporting the findings of this research paper will be made available by the authors upon a reasonable request.

REFERENCES

[1]. P. Jain, A. M. Joshi, S. P. Mohanty and L. R. Cenkeramaddi, "Non-Invasive Glucose Measurement Technologies: Recent Advancements and Future Challenges," in IEEE Access, vol. 12, pp. 61907-61936, 2024.

[2]. Y. Obeidat, "The Most Common Methods for Breath Acetone Concentration Detection: A Review," in IEEE Sensors Journal, vol. 21, no. 13, pp. 14540-14558, 1 July1, 2021.

[3]. A. K. Pathak and C. Viphavakit, "VOC Biomarker Monitoring for Diabetes Through Exhaled Breath Using Ag/P-TiO2 Composite Plasmonic Sensor," in IEEE Sensors Journal, vol. 21, no. 20, pp. 22631-22637, 15 Oct.15, 2021.

[4]. N. Alizadeh, H. Jamalabadi and F. Tavoli, "Breath Acetone Sensors as Non-Invasive Health Monitoring Systems: A Review," in IEEE Sensors Journal, vol. 20, no. 1, pp. 5-31, 1 Jan.1, 2020.

[5]. U. N. Thakur, R. Bhardwaj and A. Hazra, "A Multivariate Computational Approach With Hybrid Graphene Oxide Sensor Array for Partial Fulfillment of Breath Acetone Sensing," in IEEE Sensors Journal, vol. 22, no. 21, pp. 20207-20215, 1 Nov.1, 2022.

[6]. S. Lekha and S. M, "Recent Advancements and Future Prospects on E-Nose Sensors Technology and Machine Learning Approaches for Non-Invasive Diabetes Diagnosis: A Review," in IEEE Reviews in Biomedical Engineering, vol. 14, pp. 127-138, 2021.

[7]. P. -L. Lee, T. -M. Lee and H. -H. Chang, "Noninvasive Reflective Optical Blood Glucose Measurement System Based on Red/Green and Infrared/Green Light Amplitude Ratios," in IEEE Sensors Journal, vol. 25, no. 20, pp. 38869-38878, 15 Oct.15, 2025.

