

BREATH BASED BIO- SIGNAL ANALYSER FOR DISEASE DETECTION

Dharshini K¹, Esha R², Harsshitha S³, Vaishnavi P⁴

¹Student, School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Tamil Nadu, India.

²Student, School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Tamil Nadu, India.

³Student, School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Tamil Nadu, India.

⁴Student, School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Tamil Nadu, India.

ABSTRACT

DHEV (Detecting Harmful Exhaled Vapors) is a novel, portable, non-invasive system designed for diabetes and gastrointestinal disease detection using breath samples. It employs metal-oxide-semiconductor (MOS)-type electrochemical sensors to analyze volatile organic compounds (VOC'S) in breath, differentiating between diseased individuals and non-diseased individuals. The system integrates with sensor data, achieving a test and train accuracy of 98.00% and 98.50% through an MLP model. This affordable digital health device aims to enhance patient adherence to regular monitoring and facilitate early intervention for diabetes [1].

Keyword: Volatile Organic Compounds (VOCs), Diabetes detection, Gastrointestinal disease, Electrochemical sensors, Non-invasive diagnostics, Portable health device, Multilayer Perceptron (MLP), Early disease detection

1. INTRODUCTION

The World Health Organization's most recent statistics indicate that there are more than 500 million diabetic patients worldwide, and around 1.6 million people die each year due to diabetes and related disorders and the number of people living with diabetes may reach 780 million by 2045, according to reports from the International Diabetes Federation (IDF) [2]. Breath analysis for disease detection involves measuring volatile organic compounds (VOCs) in breath using sensitive analytical techniques. Disease diagnosis is conventionally conducted using expensive, time-consuming, invasive techniques, applied by appropriately trained health care professionals. For instance, gastroscopy, laryngoscopy, and coronary angiography are used for gastric cancer (GCa), lung cancer (LC), and myocardial infarction diagnosis, respectively [3]. Other commonly used methods, such as computed tomography [4] or mammography, used for breast cancer (BC) [5], may also be harmful due to radiation exposure. As a result, patient compliance and utilization of such diagnostic methods are remarkably reduced for a significant part of the population. In this project we use non-invasive method for disease detection by using various sensors such as MQ-3 for diabetic detection using acetone as VOC, MQ-135 for gastro intestinal disorders using ammonia as VOC. Since the gas sensors are sensitive to temperature and humidity, we use DHT22 in order to generalise the reading by setting a threshold level.

1.1 Breath analysis for Diabetes

Acetone is not only an effective biomarker of Diabetes Mellitus but also proved to be a rapid, patient compliant viable alternative to the conventional methods of blood glucose determination [6]. The correlation between specific VOCs in human breath and blood glucose levels (BGLs) has been established. For instance, breath acetone has been identified as a crucial biomarker for type 2 diabetes, directly linked to blood glucose concentrations [7,8]. Previous studies have shown that the acetone concentration in breath correlates with BGLs [8]. Similarly, a recent study [9] investigated the potential of salivary amylase as a biomarker for diabetes. Salivary amylase studies have shown that the acetone concentration in breath correlates with BGL.

1.2 Breath analysis for gastrointestinal diseases

Ammonia in breath serves as a biomarker for conditions such as chronic kidney disease (CKD), liver dysfunction, and gastrointestinal disorders like *Helicobacter pylori* infection [10]. Elevated breath ammonia levels correlate strongly with blood urea nitrogen (BUN) levels, making it useful for assessing kidney function in CKD patients. Among the many markers, ammonia is often used as a biomarker for the monitoring of liver and kidney functions [11].

2. LITERATURE REVIEWS

Kapur et al. (2023) developed *DiabeticSense*, a portable IoT-based system for non-invasive diabetes detection through breath analysis, aiming to replace conventional finger-prick tests. The system addresses infrequent glucose monitoring by providing a painless alternative, particularly beneficial for early-stage detection and resource-constrained environments, they have taken clinical samples from 100 patients at a nationally recognized hospital, they have used multi-sensor architecture with cloud connectivity for data processing, the obtained data was classified and modeled using Gradient Boosting Classifier with cross-validation they have obtained an accuracy of 86.6% [1].

The study by Paleczek, Grochala, and Rydosz (2021) Developed a non-invasive breath analysis system using acetone as a biomarker for diabetes detection through machine learning. They have acetone as the VOC of diabetic detection using a multi-sensor array where the controlled experiment has acetone concentration (0.9-5.4) and classification is done by XGBoost algorithm with 92% accuracy and Deep learning model with 96% accuracy. They created synthetic data using CTGAN and with that data they have trained and improved the model [17].

In this paper the model is trained for detecting type 2 diabetics by measuring the acetone levels. They made a hybrid approach and named it as CORNN which is the combination of SVM and MLP with different kernels to improve the model performance. The dataset comprised of 152 breath samples with 68 diabetic and 84 healthy individuals with this hybrid model. They obtained an accuracy of 98.02% with low error rate of 0.02 [2].

The article analyzes existing studies using VOCs in breath for lung cancer detection. It reviews datasets from multiple studies, including various sample sizes and gas chromatography-mass spectrometry (GC-MS) and electronic nose (e-nose) methods. Reported model accuracy varies from 71% to over 90%, depending on sensor type and VOC detection approach. However, no single unified dataset or model is established due to variability across studies [16].

This paper aims to develop a non-invasive glucose monitoring system that utilizes exhaled breath analysis, integrating IoT devices with machine learning algorithms for real-time glucose level estimation. They have collected 462 samples from individuals, they reviewed and used 13 machine learning algorithms such as random forest, SVM, XGBoost and Gradient boosting and achieved an accuracy of 96.9%, the model is trained for testing the blood sugar level. These results indicate a significant improvement over previous methods, with the BGL (Blood Glucose Level) assessment model achieving a 43.3% higher accuracy compared to earlier approaches [18].

This review paper delves into advancements within electronic nose (e-nose) technology for exhaled breath analysis. It's particularly focused on highlighting the various sensors used, with specific attention to the MQ series, in detecting VOCs relevant to different diseases. The paper likely discusses the specific design considerations and performance metrics of these sensors, offering insights into the current capabilities and limitations of e-noses in clinical applications [19].

This survey offers a broad overview of non-invasive breath analysis techniques aimed at biomarker detection for clinical diagnosis. It likely explores a range of algorithms used in the analysis process. A key aspect is the discussion of the advantages and limitations, as well as overarching challenges, in using breath analysis for early disease detection and prognosis [20].

This paper focuses specifically on the use of exhaled breath VOC biomarkers for screening liver diseases and head and neck cancers, covering both metabolic and non-metabolic disorders. It examines the application of single sensors and e-nose (sensor array) approaches, combined with machine learning techniques for pattern recognition. This research demonstrates the increasing use of data-driven methods to enhance the accuracy of breath-based diagnostics [21].

This study assesses the feasibility of using a low-cost, sensor-based system for monitoring methane levels. The key finding is that combining TGS2611 and MQ-4 sensors improves methane detection accuracy and range. This enhancement is particularly relevant for monitoring gastrointestinal disorders where methane serves as a key biomarker, showcasing the potential for affordable, non-invasive diagnostics [22].

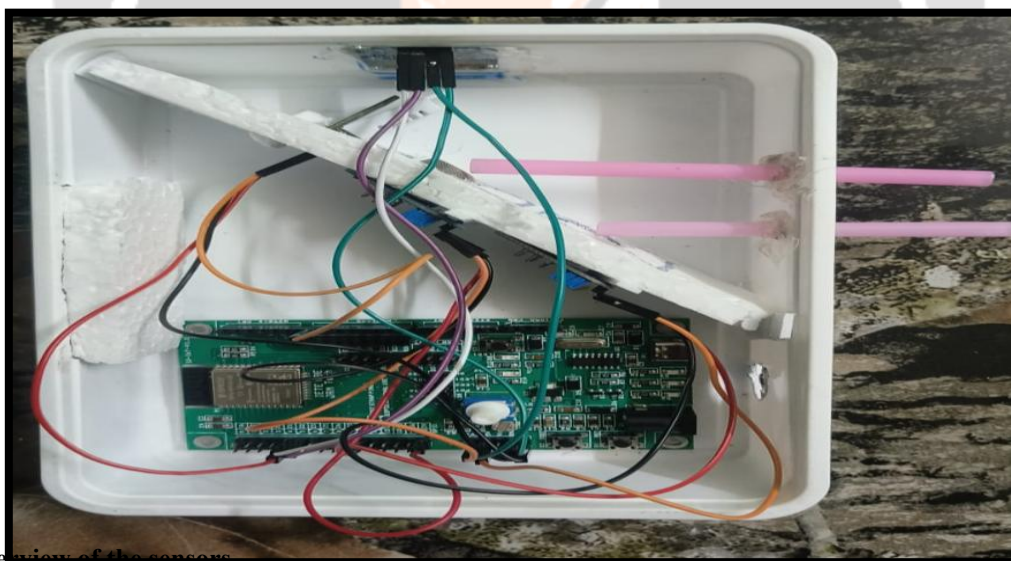
3. MATERIALS AND METHODS

3.1 Objective

To develop and evaluate a low-cost portable, non-invasive device named DHEV that uses breath samples as input and generates diabetes and gastrointestinal disease predictions based on Deep Learning model.

3.2 Design

The breath analysis enclosure is meticulously designed to ensure accurate sensor exposure and reliable data acquisition. An inclined thermocol platform is diagonally positioned within the enclosure to support the sensors, providing both elevation and stability. This orientation ensures direct exposure of the sensors to the incoming breath sample. At the front panel, two pink straws serve as dedicated breath inlet tubes, aligned precisely with the sensor array to enable efficient capture of volatile organic compounds (VOCs). The sensor assembly consists of MQ-3, MQ-135, and DHT22 sensors, strategically spaced on the platform to prevent thermal and signal interference, thereby enhancing the accuracy of measurements. The ESP32 development board is securely mounted at the base of the enclosure, with neatly organized jumper wires establishing connections to all sensors and the integrated OLED display. The compact OLED screen is used to present real-time readings of temperature, humidity, and gas concentration values, enabling immediate monitoring and validation of sensor data.



3.3 Overview of the sensors

Fig -1: DHEV Device

- **DHT 22**

DHT22 is a commonly used temperature and humidity sensor. The sensor has a dedicated NTC thermistor to measure temperature and a 12-bit microcontroller to output temperature and humidity values as serial data. The sensor can measure temperature from -40 °C to 80 °C and humidity from 0% to 100% with an accuracy of $\pm 1^\circ\text{C}$ and $\pm 1\%$ [12].

- **MQ-3 (Acetone Sensor)**

The MQ-3 is a widely used analog gas sensor designed to detect alcohol and ketones, particularly acetone, which is a key biomarker in diabetes-related breath analysis. It uses a tin dioxide (SnO_2) semiconductor layer whose resistance decreases in the presence of acetone vapors. The sensor provides an analog voltage output that varies with gas concentration. It operates effectively in the range of 0.1 to 10 ppm acetone and has high sensitivity with good stability.

- **MQ-135 (Ammonia Sensor):**

The MQ-135 is an air quality sensor capable of detecting gases such as ammonia, sulfur dioxide, carbon monoxide, and benzene. Like other MQ sensors, it relies on a tin dioxide-based sensing layer that changes resistance in response to gas concentration. It has a detection range of 10 to 1000 ppm for ammonia and provides analog voltage output proportional to concentration.

Table -1: Threshold range for disease detection

<i>DISEASE</i>	<i>NORMAL LEVEL</i>	<i>DISEASE LEVEL</i>
Diabetes (Acetone)	<1.8ppm	>1.8ppm
Gastro Intestinal (Ammonia)	<1.5ppm	>1.5ppm

Table -2: Real-time effect of fasting in biomarker

<i>TIME</i>	<i>Tested Value mg/dL</i>	<i>Value of MQ3 sensor (ppm)</i>
Before fasting	180	1.90
1/2 hour after fasting	226	2.10
1 hour after fasting	220	2.10
1 1/2 hour after fasting	205	2.00
2 hour after fasting	184	1.90

The patient who was tested for this study has type 2 diabetic with a usual range of 180-230 mg/dL. The tested value is measured using One Touch glucose level testing machine.



4. MODEL DEVELOPMENT

4.1 Architecture

A Multi-Layer Perceptron (MLP) classifier was used for disease prediction. The model architecture and training parameters were:

- Model: MLP
- Hidden Layers: Two hidden layers with 32 neurons each
- Activation Function: ReLU
- Learning Rate: Initial learning rate of 0.1
- Regularization: L2 regularization with alpha = 0.1
- Early Stopping: Enabled with validation fraction of 0.1 and patience parameter of 20 epochs
- Random State: 42 for reproducibility
- Maximum Iterations: 1500



Fig -3: DHT22 [15]



Fig -4: MQ135 [14]

Table -3: Comparision between Logistic Regression and MLP model

LOGISTIC REGRESSION	MLP MODEL
Uses linear relationship	Uses non-linear relationship
Accuracy: 20-50%	Accuracy 90-98%
Not suitable for serial values	Best suitable for serial values
Suitable for only classes with linear boundaries	MLP can learn complex, curved boundaries..
Sensitive to noise	More robust to sensor noise

4.2 Model Training

The dataset was split into training and testing sets with an 80/20 ratio. The model was trained on the training set, and its performance was evaluated on the test set. The evaluation metrics used were:

1. Accuracy
2. F1 Score
3. Precision
4. Recall
5. Confusion Matrix
6. Cross-Validation Score (5-fold stratified cross-validation)

4.3 Assessment of the evaluation metrics

- 1) Accuracy: It represents the ratio of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- 2) F1 Score: It is the harmonic mean of precision and recall

$$F1 = 2 * \frac{(\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

- 3) Precision: Proportion of correctly predicted positive observations to the total predicted positive observations.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- 4) Recall: Proportion of correctly predicted positive observations to all actual positives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where:

- **TP** = True Positives

- **TN** = True Negatives
- **FP** = False Positives
- **FN** = False Negatives

5) Confusion matrix: It is a tabular representation of actual versus predicted classifications. It illustrates the number of correct and incorrect predictions made by the model, categorized by class.

6) Cross-Validation (CV) Score (5-fold stratified cross-validation): Stratified k-fold cross-validation partitions the dataset into k equally sized folds, preserving the class distribution in each fold. Each fold is used once as a validation set while the remaining k-1 folds are used for training.

$$CV \text{ Score} = \frac{1}{k} \sum_{i=1}^k \text{Score}$$

Table -4: Hyper parameter tuning table

S.No	Hyperparameter Tuning				
	Hidden layers	Alpha	Learning rate	F1 score- Training	F1 score- Testing
1	(32)	0.0001	0.001	78.4%	76.0%
2	(64)	0.001	0.001	82.1%	85.1%
3	(64, 32)	0.01	0.01	88.6%	94.7%
4	(64, 32, 16)	0.05	0.01	91.3%	94.1%
5	(128, 64, 32)	0.05	0.005	90.7%	93.8%
6	(64, 64)	0.01	0.01	89.5%	92.5%
7	(32, 32, 32)	0.01	0.001	87.9%	83.0%
8	(128)	0.001	0.01	83.2%	81.2%
9	(128, 64)	0.001	0.005	88.0%	89.5%
10	(64, 32, 16)	0.001	0.005	87.3%	90.7%
11	(32, 32)	0.1	0.1	98.5%	97.9%

4.4 Dataset

We generated a based on the from the and combined it time data the breath diabetic patients.

collection

synthetic dataset threshold values research papers with the real collected from samples from 8

5. RESULT

The developed Multi-Layer Perceptron (MLP) model exhibited outstanding performance in classifying diabetes and gastrointestinal diseases based on breath biomarkers. The model achieved a training accuracy of 98.50% and a testing accuracy of 98.00%. The weighted F1 scores 0.9850 for training and 0.9799 for testing further confirmed that the model maintained a strong balance between precision and recall across all disease categories. Cross-validation using a stratified 5-fold approach yielded a mean F1 score of 0.9312 ± 0.0370 , indicating consistent performance across different data subsets.

6. CONCLUSION AND FUTURE SCOPE

The DHEV system successfully bridges the gap between traditional invasive diagnostics and modern breath-based disease detection. By integrating ESP-32 with MQ-3 (acetone) and MQ-135 (ammonia) sensors developed with an optimized MLP model, the system provides a portable, non-invasive, and cost-effective solution for early disease screening. The high accuracy (98%) and strong F1 scores validate its clinical potential, particularly for diabetes and gastrointestinal disorders, where breath biomarkers like acetone and ammonia have established diagnostic relevance. The current breath-based disease detection system demonstrates promising results for early screening of diabetes and gastrointestinal disorders. However, it can be improved by using more specific gas sensors to detect a wider range of diseases more accurately. Collecting additional real-time data

from different people and settings will help make the model more reliable. An interface can be developed to show live breath readings, disease predictions, and past trends. Also, adding wireless features like Bluetooth or Wi-Fi will allow the system to send data to a mobile phone or computer, making it easier to use for regular health checks, even in remote areas.

7. ACKNOWLEDGEMENT

We sincerely thank our Dean, Dr. Soman K. P., for providing us with the opportunity to undertake this project. We extend our gratitude to our course coordinator, Dr. Kelath Murali Manoj, and our professors, Dr. Amrutha V. and Dr. Snigdhatanu Acharya, for their guidance and support. We also acknowledge all individuals who contributed to the successful completion of our project.

8. REFERENCES

- [1]. Kapur, R., Kumar, Y., Sharma, S., Rastogi, V., Sharma, S., Kanwar, V., Sharma, T., Bhavsar, A., & Dutt, V. (2023). DiabeticSense: A Non-Invasive, Multi-Sensor, IoT-Based Pre-Diagnostic System for Diabetes Detection Using Breath. *Journal of Clinical Medicine*, 12(20),6439. <https://doi.org/10.3390/jcm12206439>
- [2]. Bhaskar, N., Bairagi, V., Boonchieng, E., & Munot, M. V. (2023). Automated detection of diabetes from exhaled human breath using deep hybrid architecture. *IEEE access*, 11, 51712-51722.
- [3]. Chen, T., Liu, T., Li, T., Zhao, H., & Chen, Q. (2021). Exhaled breath analysis in disease detection. *Clinica chimica acta; international journal of clinical chemistry*, 515,6172. <https://doi.org/10.1016/j.cca.2020.12.036>
- [4]. Gashimova, E., Temerdashev, A., Porkhanov, V., Polyakov, I., Perunov, D., Azaryan, A., & Dmitrieva, E. (2020). Investigation of different approaches for exhaled breath and tumor tissue analyses to identify lung cancer biomarkers. *Heliyon*, 6(6), e04224. <https://doi.org/10.1016/j.heliyon.2020.e04224>
- [5]. Phillips, M., Cataneo, R. N., Saunders, C., Hope, P., Schmitt, P., & Wai, J. (2010). Volatile biomarkers in the breath of women with breast cancer. *Journal of breath research*, 4(2), 026003. <https://doi.org/10.1088/1752-7155/4/2/026003>
- [6]. Sudi, R., Biligere, V., R, T., P, T. S., & M, T. (2021). Analysis of Volatile Organic Compounds in Exhaled Breath for Detection of Diabetes Mellitus. <https://doi.org/10.1109/GCAT52182.2021.9587790>
- [7]. Anderson J. C. (2015). Measuring breath acetone for monitoring fat loss: Review. *Obesity* (Silver Spring, Md.), 23(12),2327–2334. <https://doi.org/10.1002/oby.21242>
- [8]. Sun, M., Chen, Z., Gong,Z.,(2015). Determination of breath acetone in 149 type 2 diabetic patients using a ringdown breath-acetone analyzer. *Analytical and bioanalytical, chemistry*, 407(6), 1641-1650. <https://doi.org/10.1007/s00216-014-8401-8>
- [9]. Pérez-Ros, P., Navarro-Flores, E., Julián-Rochina, I., Martínez-Arnau, F. M., & Cauli, O. (2021). Changes in Salivary Amylase and Glucose in Diabetes: A Scoping Review. *Diagnostics*, 11(3),453. <https://doi.org/10.3390/diagnostics11030453>
- [10]. Chan, M. J., Li, Y. J., Wu, C. C., Lee, Y. C., Zan, H. W.,Meng, H. F., Hsieh, M. H., Lai, C. S., & Tian, Y. C. (2020). Breath Ammonia Is a Useful Biomarker Predicting Kidney Function in Chronic Kidney Disease Patients. *Biomedicines*, 8(11),468. <https://doi.org/10.3390/biomedicines8110468>
- [11]. Rath, R. J., Herrington, J. O., Adeel, M., Güder, F., Dehghani, F., & Farajikhah, S. (2024). Ammonia

- detection: A pathway towards potential point-of-care diagnostics. *Biosensors & bioelectronics*, 251, 116100.
<https://doi.org/10.1016/j.bios.2024.116100>
- [12]. <https://www.kuongshun-ks.com/uno/uno-sensor/dht22-am2302-digital-temperature-humidity-sensor.html>
- [13]. <https://www.electroniccomp.com/mq3-alcohol-gas-sensormoduleindia?srsId=AfmBOqXEZm3om7UX1H1CCpGH5kkX3qXNPYZKeALJxIKKXSYYKA7d8NKGAg>
- [14]. Mardikar, S. (2015). Detection of ammonia in exhaled human breath. *IOSR Journal of Electronics and Communication Engineering*, 3(4), 119-27.
- [15]. <https://esp32tutorials.com/dht22-esp32-esp-idf/>
- [16]. Manoj Khokhar 2024 J.Breath Res. 18 024001 DOI 10.1088/1752-7163/ad15fb
<https://iopscience.iop.org/article/10.1088/1752-7163/ad15fb/pdf>
- [17]. Anna Paleczek, Dominik Grochala, Artur Rydosz, "Artificial Breath Classification Using XGBoost Algorithm for Diabetes Detection", 2021-06-18.
[10.3390/s21124187](https://doi.org/10.3390/s21124187)
- [18]. R. Kapur et al., "GlucoBreath: An IoT, ML, and Breath-Based Non-Invasive Glucose Meter," in *IEEE Access*, vol. 12, pp. 59346-59360, 2024, doi: 10.1109/ACCESS.2024.3392015.
<https://ieeexplore.ieee.org/abstract/document/10506520>
- [19]. Li et al. (2023), "Research progress of electronic nose technology in exhaled breath disease analysis." *Microsystem Technology* **9**, 129 (2023).
<https://doi.org/10.1038/s41378-023-00594-0>
- [20]. Mule (2021), "A comprehensive survey on investigation techniques of exhaled breath (EB) for diagnosis of diseases in human body." *Informatics in Medicine Unlocked*, Volume 26, 2021
<https://doi.org/10.1016/j.imu.2021.100715>
- [21]. Hripriya P et al. (2023), "Breath VOC analysis and machine learning approaches for disease screening." 2023 J. Breath Res. **17** 024001, volume 17.
<https://iopscience.iop.org/journal/1752-7163>
- [22]. Nagahage et al. (2021), "Assessment of the applicability of a low-cost sensor-based methane monitoring system for continuous multi-channel sampling." *Environ Monit Assess* **193**, 509 (2021).
<https://doi.org/10.1007/s10661-021-09290-w>