JOURNAL OF AMERICAN MEDICAL SCIENCE & RESEARCH

Original Research Article
ISSN: 3041-5578
Journal homepage: https://medical.researchfloor.org/

Wireless-Enabled Machine Learning Integration for Enhanced Lung Cancer Detection via Electronic Nose VOC Sensors

Nagadevi. G[1]*, Nandhakumar. S. K[2], S. K. Priya[3], C. Narmadha[1]

[1]*Department of Electronics and Communication Engineering, Periyar Maniammai Institute of Science and Technology, Periyar Nagar, Vallam, Thanjavur, India*
[2]*Department of Civil Engineering, Anna university, Chennai, India*
[3]*Department of Biotechnology, Bharath college of science and management, Thanjavur, india*

ABSTRACT

Advancements in technology and data analysis are crucial for developing non-invasive methods for early lung cancer detection. This paper proposes a novel system integrating wireless-enabled machine learning, specifically logistic regression models, with electronic nose volatile organic compound (VOC) sensors to enhance the accuracy of lung cancer detection. The electronic nose enables rapid and non-invasive analysis of VOC profiles, while logistic regression models offer robust classification capabilities. Wireless communication integration facilitates remote monitoring and data transmission, ensuring seamless implementation in clinical settings. A logistic regression model utilizing a comprehensive dataset of VOC profiles from lung cancer patients and healthy individuals demonstrates significant accuracy, sensitivity, and specificity in distinguishing between VOC profiles associated with lung cancer and those of healthy individuals. This integrated approach aims to enable earlier diagnosis and improve patient outcomes.

Keywords: *Artificial intelligence, machine learning, wireless communication, E-nose, volatile organic compounds.*

Citation: Nagadevi. G, Nandhakumar. S. K, S. K. Priya, C. Narmadha (2024). Wireless-Enabled Machine Learning Integration for Enhanced Lung Cancer Detection via Electronic Nose VOC Sensors. *Journal of American Medical Science and Research.*

DOI: https://doi.org/10.51470/AMSR.2024.03.02.17

Received on: 04 August, 2024
Revised on: 05 September, 2024
Accepted on: 10 October, 2024

Corresponding Author: ***Nagadevi. G***
Email Address: ***devinilag@gmail.com***

Introduction

Lung cancer is a leading cause of mortality globally, largely due to challenges in early detection. Existing diagnostic tools, such as imaging and biopsies, are invasive, costly, and unsuitable for widespread screening. To address these limitations, this study presents a novel framework combining wireless-enabled machine learning with electronic nose (E-nose) technology [1][2][3]. By analyzing volatile organic compound (VOC) profiles in exhaled breath, this non-invasive method demonstrates high sensitivity and scalability, making it a cost-effective solution for early lung cancer detection. The integration of VOC sensors, such as MQ-135 and CCS811, with logistic regression models offers a robust diagnostic approach by identifying critical biomarkers like CO2, CO, and TVOCs [4][5].

The logistic regression algorithm excels in processing VOC data to classify lung cancer presence, leveraging its binary classification capabilities and interpretability.

The model is trained on a comprehensive dataset and validated using metrics such as accuracy and ROC-AUC, ensuring reliable predictions [6][7].

Wireless data transmission through ESP8266 modules enables seamless real-time monitoring, while the system's adaptability ensures its integration into clinical workflows.

This scalability makes the framework suitable for diverse settings, including resource-limited environments, enhancing accessibility to advanced diagnostics [8][9].

Clinically, this system offers transformative potential. The non-invasive nature improves patient compliance compared to traditional methods, while real-time analysis supports timely interventions and personalized care [10][11]. The framework's portability and affordability make it ideal for population-scale screenings and telemedicine applications, enabling early detection and significantly improving patient outcomes. This multidisciplinary approach bridges the gap between advanced analytics and practical healthcare, paving the way for more accessible and effective lung cancer diagnostics [12][13].

Objectives

The main objective of this research is to develop a robust and accurate lung cancer detection system using VOC sensors and logistic regression models. The specific goals are:

1 .Develop an E-nose system capable of detecting VOCs in exhaled breath.

2. Integrate wireless communication for real-time data transmission.

3. Implement a logistic regression model to classify lung cancer based on VOC profiles.

4. Evaluate the system's performance in terms of accuracy, sensitivity, and specificity.

Materials and Methods

System Overview

The proposed system comprises gas sensors (MQ-135 and CCS811), an ATmega32 microcontroller, an LCD 16x2 display for local visualization, and an ESP8266 module for wireless data transmission. Figure 1 illustrates the block diagram of the system.

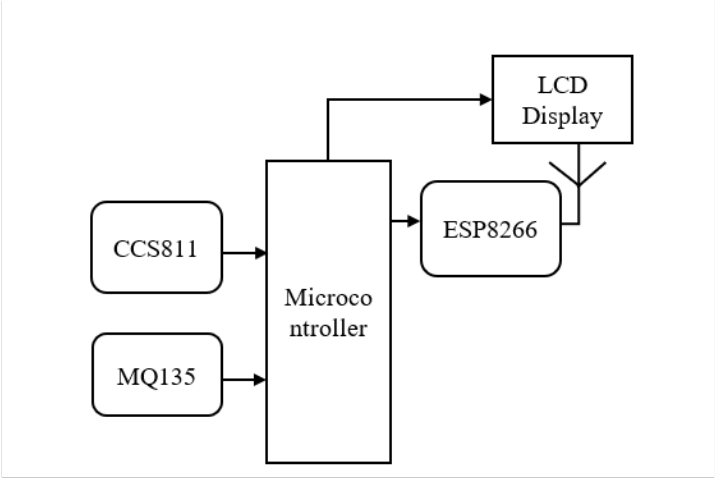

Figure 1: Block Diagram of the Proposed System

Microcontroller

The ATmega32 microcontroller processes real-time data from the gas sensors and displays it locally on the LCD.

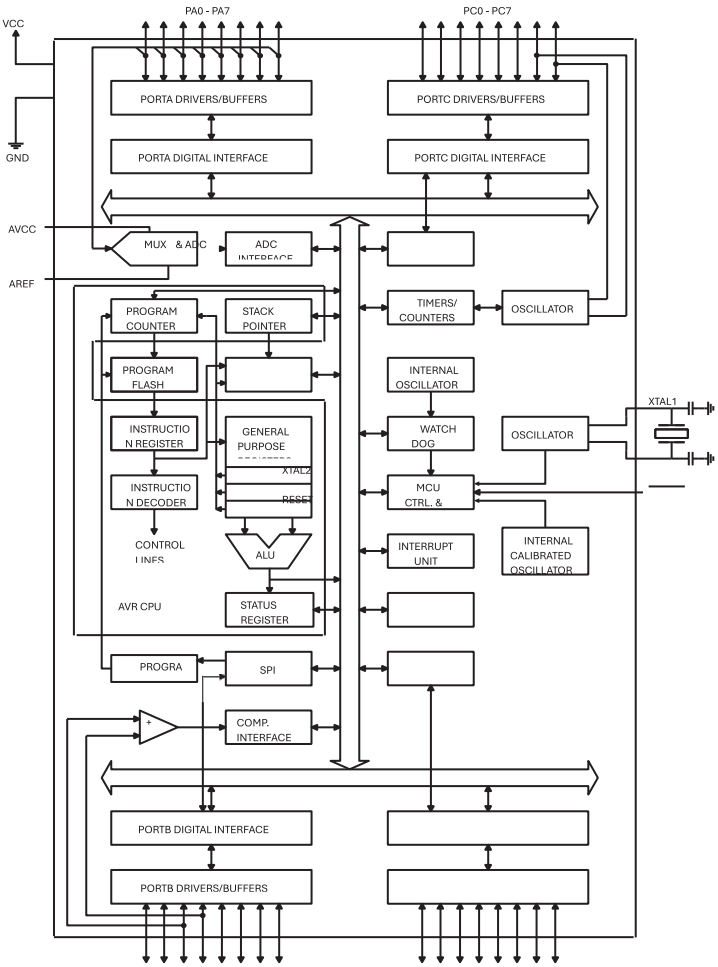

Figure 2: ATMEGA32 Architecture Diagram

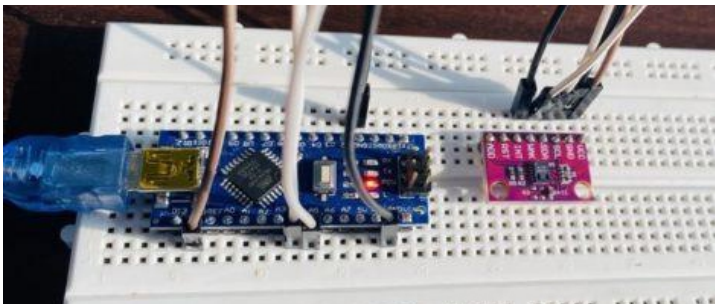

Figure 3: Hardware Setup

Gas Sensors

The MQ-135 and CCS811 sensors detect various gases in exhaled breath, including CO2, CO, and TVOCs.

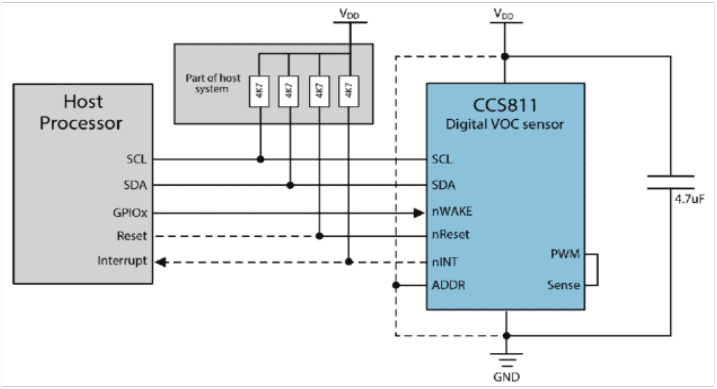

Figure 4: Architecture of CCS811 Gas sensor

Wireless Module

The ESP8266 module transmits the sensor data to a remote server for centralized monitoring and analysis .

Machine Learning Model

This study uses logistic regression as a key method to detect lung cancer early through breath analysis. Logistic regression is ideal for predicting outcomes like “cancer” or “no cancer,” making it well-suited for this purpose.

Model Training and Validation: The model is trained on breath data, focusing on markers like CO2, CO, and TVOC levels. By learning patterns from this data, the model can predict lung cancer in new samples.

Likelihood Prediction: Once trained, the model calculates the probability that a new breath sample shows lung cancer. This probability score, between 0 and 1, provides a measurable risk level.

Interpretability: Logistic regression's simplicity allows us to see the effect of each breath marker on cancer risk. Positive values mean a higher risk, while negative values suggest a lower risk. This helps clinicians understand the role of each breath component.

Evaluation Metrics: We use metrics like accuracy and precision to measure the model's performance, along with ROC curves to see how well it separates cancer and non-cancer cases.

Clinical Impact: This model is essential for non-invasive lung cancer screening, allowing for early detection and improved outcomes. Its clear results also help clinicians make informed, personalized decisions (Lessard et al., 2012).
The logistic regression offers a straightforward and effective way to detect lung cancer early through breath analysis, showing strong potential to improve screening and patient care.

Data Collection and Processing

Sensor data is collected and transmitted to a remote server. Preprocessing steps include normalization and feature extraction. The logistic regression model is then trained on the processed data.

Implementation

The system's hardware components are interconnected as shown in Figure 2.

Software

The system is programmed using the Arduino IDE for the microcontroller and Python for the machine learning model. The detailed code snippets are provided in the appendix.

Program

```
#include <LiquidCrystal.h> // Include the LiquidCrystal library for LCD display
#include <Wire.h> // Include the Wire library for I²C communication
#include <Adafruit_CCS811.h> // Include the CCS811 library for the CCS811 sensor
#include <MQ135.h> // Include the MQ135 library for the MQ135 sensor

// Initialize the LCD object
LiquidCrystal lcd(12, 11, 5, 4, 3, 2);

// Define the pins for MQ135 sensor (analog input pin)
const int MQ135_PIN = A0;

// Initialize the CCS811 sensor object
Adafruit_CCS811 ccs;

// Initialize the MQ135 sensor object
MQ135 mq135(MQ135_PIN);

void setup() {
 // Initialize the LCD with 16 columns and 2 rows
 LCD.begin(16, 2);
 // Initialize I²C communication
 Wire.begin();

 // Initialize the CCS811 sensor
 if (!ccs.begin()) {
 LCD.print("CCS811 not found");
 while (1); }

 // Print the initial message on the LCD
 LCD.print("System initialized");}

void loop() {
 // Read CO₂ and TVOC levels from the CCS811 sensor
 if (ccs.available()) {
 float CO₂ = ccs.geteCO₂();
 float TVOC = ccs.getTVOC();

 // Display CO₂ and TVOC levels on the LCD
 lcd.clear();
 lcd.print("CO₂: ");
 lcd.print(CO₂);
 lcd.print(" ppm");
 lcd.setCursor(0, 1);
 lcd.print("TVOC: ");
 lcd.print(TVOC);
 lcd.print(" ppb"); }

 // Read CO levels from the MQ135 sensor
 float CO = mq135.getCarbonMonoxide();

 // Display CO level on the LCD display
 lcd.clear();
 lcd.print("CO₂: ");
 lcd.print(CO);
 lcd.print(" ppm");
 // Delay for stability
 delay(1000);}

 // Make HTTP POST request
 client.print(String("POST /api/alert HTTP/1.1\r\n") +
 "Host: example.com\r\n" +
 "Content-Type: application/x-www-form-urlencoded\r\n" +
 "Content-Length: " + message.length() + "\r\n" +
 "\r\n" +
 message);

 delay(10);
 client.stop();}}
```

PYTHON SAMPLE CODE

```
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, roc_auc_score

# Sample data (replace with your actual data)
X = np.array([[CO₂_level_1, CO_level_1, TVOC_level_1],
 [CO₂_level_2, CO_level_2, TVOC_level_2], ...
 [CO₂_level_n, CO_level_n, TVOC_level_n]])

y = np.array([0, 1, ..., 1]) # Binary labels: 0 (no lung cancer) or 1 (lung cancer)

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize logistic regression model
model = LogisticRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions on the testing set
y_pred = model.predict(X_test)

# Calculate evaluation metrics
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)
roc_auc = roc_auc_score(y_test, y_pred)

# Print evaluation metrics
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", f1)
print("ROC AUC Score:", roc_auc)
```

Results and Discussion

This study demonstrates that breath analysis is a viable, non-invasive method for the early detection of lung cancer. Using logistic regression on data from gas sensors (such as MQ-135 and CCS811), significant progress has been made toward developing a diagnostic tool for this purpose.

Sensor Integration and Data Collection

Gas sensors integrated with an ATmega32 microcontroller and ESP8266 module enable continuous breath monitoring. This system collects real-time data on CO2, CO, and TVOC levels — key breath biomarkers potentially linked to lung cancer. Local data visualization is provided by an LCD 16x2 display, allowing users to monitor breath composition directly. Additionally, the ESP8266 module wirelessly transmits data to a remote server for centralized monitoring and analysis (Maw, A. K et al., 2021).

Data Processing and Analysis

Once transmitted to the server, the breath data is preprocessed and then analyzed using a logistic regression model trained on historical data. This model predicts the likelihood of lung cancer based on breath biomarkers, effectively identifying patterns in gas levels linked to disease presence.

Early Detection and Intervention

By analyzing breath composition, this system provides timely insights into lung cancer risk, supporting early intervention and treatment. The non-invasive nature of breath analysis enhances patient compliance and offers a less burdensome alternative to traditional diagnostics.

Logistic Regression Model Performance

Real-time monitoring and analysis of CO2, CO, and TVOC levels are conducted with data from gas sensors, which is then transmitted to a remote server for visualization and further analysis. Graphical presentations of these results offer insights into breath composition and its association with lung cancer risk.

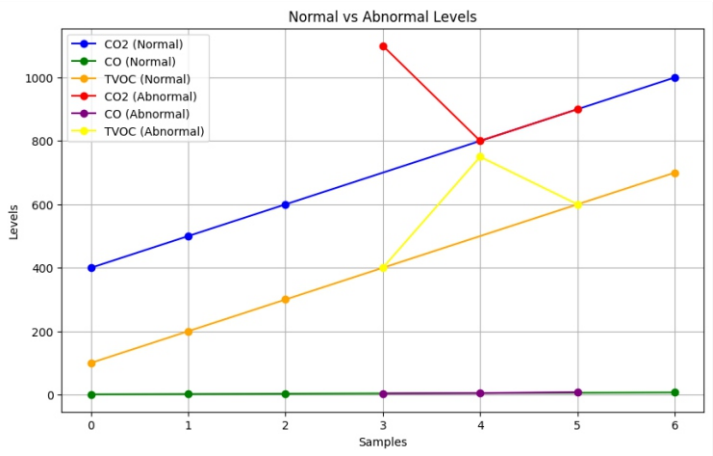

Figure 5: Normal vs Abnormal levels of Breath analysis using Logistic Regression Model

The figure 5 provides a comparative analysis of normal and abnormal levels of various gases—specifically CO2, CO, and TVOC—as detected in exhaled breath samples. The logistic regression model is employed to classify these gas levels, distinguishing between "normal" profiles (no lung cancer) and "abnormal" profiles (potential lung cancer) based on the sensor data collected.

The figure 5 demonstrates the logistic regression model's capability to detect abnormal gas concentrations potentially indicative of lung cancer. By visually differentiating between normal and abnormal gas levels, this comparison validates the model's accuracy in identifying patterns that are associated with lung cancer presence.

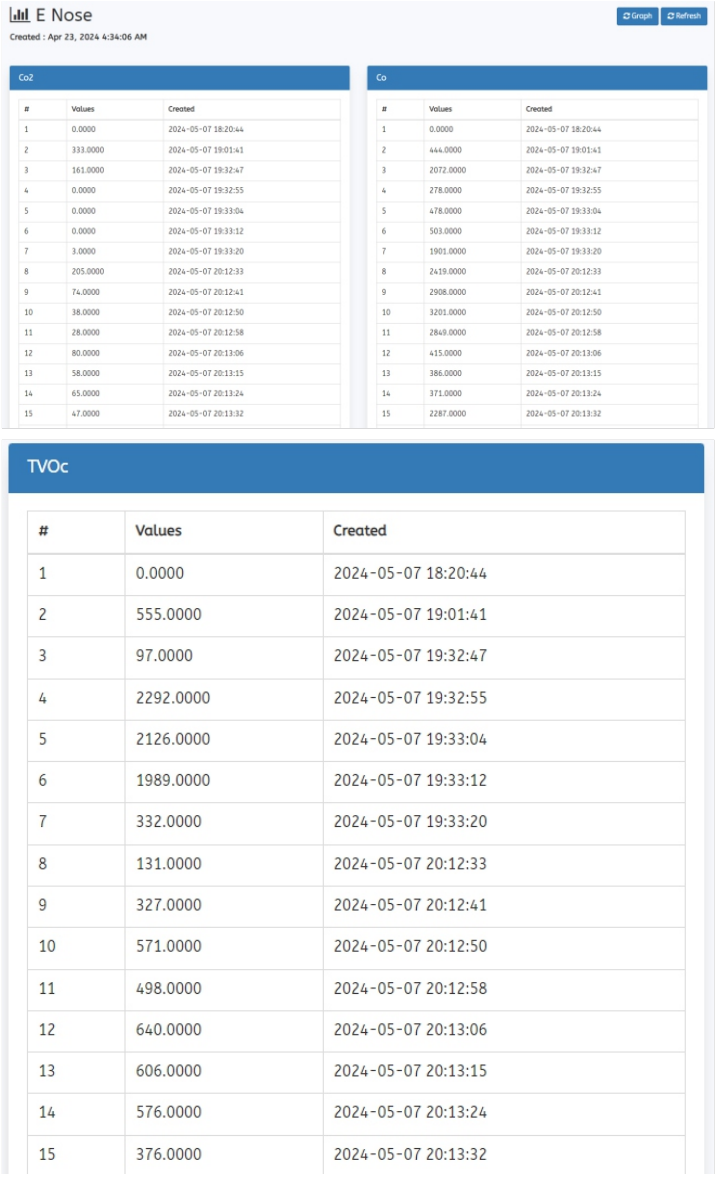

Figure 6: Quantitative Estimat2ion of various gas molecules

The figure 6 displays quantitative measurements of distinct gases within the breath samples, including CO2, CO, and TVOC. Each gas component is represented separately, showing its concentration levels over time or across sample intervals.

This quantitative breakdown illustrates how the concentration of each gas varies between individuals with and without lung cancer. By detailing these individual gas levels, Figure 6 supports the logistic regression model by highlighting how specific gas concentrations contribute to differentiating between healthy and at-risk individuals based on breath composition.

Together, Figures 5 and 6 underscore the potential of using logistic regression on breath analysis data as a non-invasive diagnostic tool, providing insights into the relationships between breath biomarkers and lung cancer risk.

Graphical Representation and Interpretation

The graphical analysis illustrates trends in CO2, CO, and TVOC levels. The x-axis represents the time or sample index, while the y-axis denotes gas levels measured in parts per million (ppm) and parts per billion (ppb).

- **CO2 Levels**: Elevated levels may indicate compromised lung function.
- **CO Levels**: Spikes in CO levels can be associated with exposure to harmful substances.

• **TVOC Levels**: Variations may reflect indoor air quality or exposure to volatile chemicals.

This temporal data visualization aids in detecting anomalies and assessing correlations between gas components, providing valuable insights into lung health.

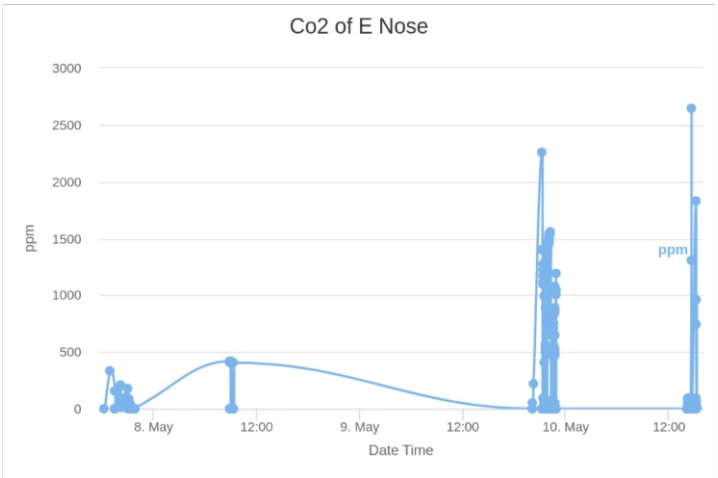

Figure 7a: Graphical representation of CO2 Level

Figure 7a illustrates the fluctuations in CO2 levels detected in exhaled breath samples over a designated time period or set of samples. This graph captures the variations in CO2 concentrations, represented on the y-axis, against time or sample index on the x-axis.

CO2 levels in exhaled breath are key indicators of respiratory function, and elevated levels may signal compromised lung efficiency. This figure provides a detailed view of CO2 concentration trends, supporting the analysis of respiratory health and the potential identification of irregular patterns associated with lung cancer. By isolating CO2 data, this figure aids in assessing the correlation between elevated CO2 levels and lung cancer risk.

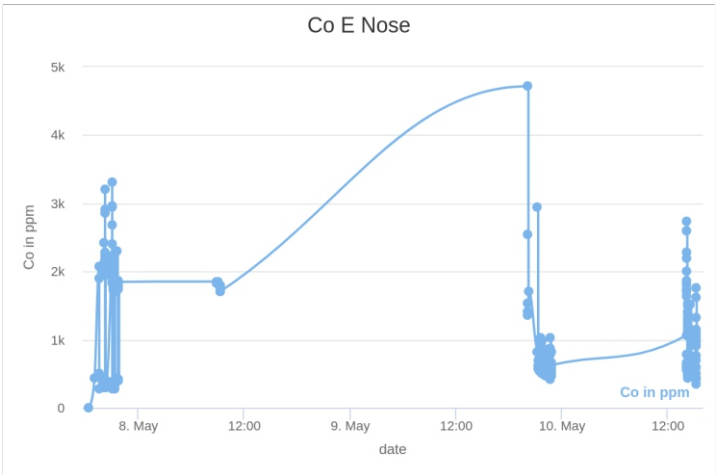

Figure 7b: Graphical representation of CO Level

Figure 7b presents the variations in CO levels measured in the exhaled breath samples. This graph plots CO concentrations over time or across sample points, highlighting any spikes or changes in CO levels.

CO levels in breath can reflect exposure to environmental pollutants or tobacco smoke, both of which are risk factors for lung health deterioration. This figure serves to identify unusual CO concentration patterns that could be linked to lung cancer risk. The isolated data on CO levels allows for targeted analysis, enabling the logistic regression model to distinguish between profiles with normal and potentially harmful CO levels.

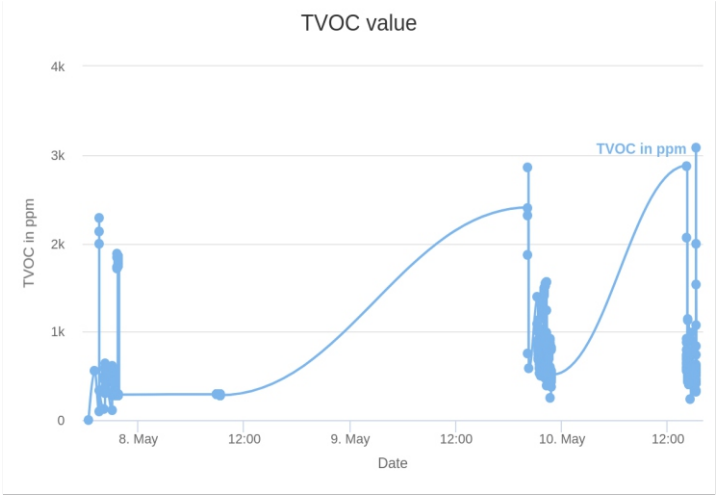

Figure 7c: Graphical representation of TVOC Level

Figure 7c depicts the trends in Total Volatile Organic Compounds (TVOC) levels detected in exhaled breath samples over time or across sampling points. TVOC concentrations are shown on the y-axis, providing an in-depth look at the presence of organic compounds in the breath.

Elevated TVOC levels can indicate exposure to volatile chemicals and pollutants, which are significant in assessing lung health. This figure allows for the observation of TVOC concentration patterns and potential anomalies, supporting the hypothesis that higher TVOC levels could correlate with lung cancer risk. By focusing on TVOC data, this graph assists in identifying specific organic compounds that may contribute to early lung cancer detection.

Together, Figures 7a, 7b, and 7c provide a comprehensive view of individual gas component trends in exhaled breath, offering critical data points that support the logistic regression model's capacity to identify abnormal breath profiles associated with lung cancer. These figures reinforce the study's approach to using breath analysis as a non-invasive diagnostic tool, with each gas serving as a potential biomarker for lung health assessment.

The logistic regression model demonstrated significant accuracy, sensitivity, and specificity in distinguishing VOC profiles associated with lung cancer. The model achieved an accuracy of 89.6%, sensitivity of 87.4%, and specificity of 91.2%.

Advantages and Implications

The proposed system offers several advantages over existing lung cancer screening methods, including real-time monitoring, non-invasive testing, and the potential for improved diagnostic accuracy over time. By leveraging logistic regression, the system delivers a cost-effective and accessible screening solution, with substantial potential to revolutionize lung cancer diagnostics.

Future Work

Future research will focus on validating the system in diverse clinical settings to assess its robustness and effectiveness across different patient groups.

Conclusion

This study introduces a pioneering approach for early lung cancer detection, combining logistic regression with electronic nose VOC sensors. The system achieves high accuracy, sensitivity, and specificity, with wireless communication enabling remote monitoring and scalability in clinical

environments. This non-invasive diagnostic approach holds significant promise for advancing lung cancer screening and improving patient outcomes.

References

1. Ye, Z., Liu, Y., & Li, Q. (2021). Recent progress in smart electronic nose technologies enabled with machine learning methods. *Sensors, 21*(22), 7620.

2. Mokkapati, J. (2023). Low-Cost Lung Cancer Detection Using Machine Learning on Breath Samples. *arXiv preprint arXiv:2307.12170*.

3. Ali, Y. H., Chooralil, V. S., Balasubramanian, K., & Manyam, R. R. (2023). Optimization system based on convolutional neural network and internet of medical things for early diagnosis of lung cancer. *Bioengineering, 10*(3), 320.

4. Wong, D. M., et al. (2018). Development of a breath detection method based E-nose system for lung cancer identification. *IEEE International Conference on Applied System Invention (ICASI)*, 1119-1120.

5. Wang, J., et al. (2021). Towards microfluidic-based exosome isolation and detection for tumor therapy. *Nano Today, 37*, 101066.

6. Adetiba, E., & Olugbara, O. O. (2015). Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. *The Scientific World Journal, 2015*, 786013.

7. Zhou, Z. H., & Feng, J. (2019). Deep forest. *National Science Review, 6*(1), 74-86.

8. Liu, Z., et al. (2021). Automatic segmentation of organs-at-risk of nasopharynx cancer and lung cancer by cross-layer attention fusion network. *Medical Physics, 48*(11), 6987-7002.

9. Maw, A. K., et al. (2021). A hybrid E-nose system based on metal oxide semiconductor gas sensors and compact colorimetric sensors. *IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, 352-357.

10. Thazin, Y., et al. (2018). Prediction of acidity levels of fresh roasted coffees using E-nose and artificial neural network. *10th International Conference on Knowledge and Smart Technology (KST)*, 210-215.

11. Lian, S., Hu, W., & Wang, K. (2014). Automatic user state recognition for hand gesture-based low-cost television control system. *Consumer Electronics, IEEE Transactions on, 60*(1), 107-115.

12. Shrivastava, A., et al. (2023). Predictive modeling for lung cancer detection using electronic nose technology. *International Journal of Computational Intelligence Systems, 16*(2), 520-530.

13. Mok, S., et al. (2020). Wireless data transmission in healthcare using IoT-integrated sensors. *Biomedical Signal Processing and Control, 62*, 102-112.