Article

Regression analysis and classification of temperature modulated metal oxide semiconductor gas sensors responses on flue gas

Elvina Chu Qing Heng, Nikko Leo, Ting Soon Ling, Hong Siang Chua, Hui En Lee*

Swinburne University of Technology Sarawak Campus

ARTICLE INFO

ABSTRACT

Industrial emissions, particularly from flue gases, pose significant risks to environmental sustainability and public health. Conventional air quality monitoring systems often suffer from high costs, delayed reporting, and limited detection capabilities. This study presents a cost-effective, real-time air quality monitoring solution using an electronic nose (eNose) system integrated with Metal Oxide Semiconductor (MOS) gas sensors. These sensors target key pollutants, such as carbon monoxide (CO) and carbon dioxide (CO_2), which also serve as indicators of transformer faults in industrial settings. The eNose system leverages machine learning for both regression and classification tasks, enabling accurate quantification of pollutant levels and categorization of air quality into defined categories. Principal Component Analysis (PCA) is employed to optimize feature extraction, enhancing model precision and efficiency. Notably, the system integrates digitally controlled buck converters for automatic temperature regulation, reducing sampling time from 390 to 130 seconds. Additionally, a redesigned airtight sensor chamber and optimized airflow design, along with the use of Tedlar bags, improve sample integrity and minimize interference. Hardware development involved prototyping on breadboards using LM2575, LM2576, and LM2574 ICs, followed by the creation of a compact 10 cm × 10 cm PCB for efficient power management. Multimeter testing verified reliable electrical connections. Experimental validation showed the system achieved over 91% accuracy in distinguishing between "good" and "bad" air quality levels. Strong correlations between sensor output and pollutant concentrations confirm system reliability. This research demonstrates a scalable, efficient tool for real-time air quality monitoring and fault detection in industrial environments.

1. Introduction

Air pollution poses a critical global challenge, significantly affecting public health and environmental sustainability. Industrial processes—especially the combustion of fossil fuels—emit flue gases containing harmful pollutants, including nitrogen oxides (NOx), sulfur oxides (SOx), carbon monoxide (CO), carbon dioxide (CO_2), volatile organic compounds (VOCs), and particulate matter. These pollutants degrade air quality, contribute to climate change, and lead to severe health risks [1]. The World Health Organization (WHO) estimates that outdoor air pollution is linked to approximately 4.2 million deaths annually, underscoring the urgent need for accurate, real-time air quality monitoring systems [2]. In response to these challenges, this study employs the Metal Oxide Semiconductor (MOS) gas sensor module previously developed in our earlier work [3–5]. In that study, temperature modulation was applied to the MOS sensors by varying the heater voltage to influence each sensor's gas selectivity, also referred to as its sensitivity profile. This technique significantly increased the dimensionality of sensor data. By exploring the multi-dimensional nature of the generated data, this study aims to identify the optimal heater voltage setting that enables precise selectivity toward specific

flue gas components. Flue gas, also referred to as exhaust gas, is the byproduct of burning fossil fuels like coal, oil, or gas in industrial applications such as boilers and engines [6]. These gases, mainly composed of CO_2 and CO, are generated from sectors such as energy, transportation, and manufacturing [2]. While these activities are essential for industrial operations, they contribute substantially to air pollution and climate change. In Malaysia, this issue has become increasingly serious, especially in urban areas with intense industrial activity and heavy traffic. Rapid urbanization has worsened the situation by increasing industrial production and vehicle emissions, further deteriorating air quality. In addition to air pollution, transformer faults are a growing concern. Faults such as overheating and insulation breakdowns can lead to the release of CO_2 and CO into transformer oil. These issues can result in expensive equipment damage, unplanned operational downtime, and disruptions to the power supply. Effective monitoring of these gases is crucial for early fault detection in transformers and to control pollution levels in urban areas. Traditional gas monitoring technologies like Continuous Emission Monitoring Systems (CEMS), Gas Chromatography (GC), and Mass Spectrometry (MS) provide precise results but are often expensive, complex, and unsuitable for large-scale deployment due to their reliance on specialized equipment and trained personnel. Furthermore, these methods typically lack real-time capabilities, making them inefficient for continuous air quality assessment [7].

To address these limitations, this study explores the development of an improved Metal Oxide Semiconductor (MOS)-based gas sensor system, commonly referred to as an electronic nose (eNose). These sensors mimic the human olfactory system, offering a cost-effective, compact, and scalable solution for detecting gas compositions, such as CO and CO_2, in both flue gases and transformer environments. Integrated with regression analysis, Principal Component Analysis (PCA), and machine learning classification algorithms, the eNose can accurately quantify pollutant concentrations and classify air quality levels in real time [8,9]. Despite their potential, MOS sensors face challenges, particularly with manual temperature modulation. This process involves adjusting the sensor's temperature to improve selectivity and sensitivity. However, manual control often results in delays between temperature adjustments and sensor response, reducing measurement reliability [10,11]. Inconsistent readings due to air leaks in non-airtight sensor chambers and time-consuming sample collection methods further compromise accuracy and efficiency. Automating the temperature modulation process can enhance sensor precision and consistency, enabling more reliable gas differentiation in complex mixtures like flue gases. Additionally, improving air sampling methods to reduce collection time and ensure airtight conditions will further streamline the system's performance. Ultimately, this study aims to design a MOS-based gas sensor system that delivers accurate, efficient, and real-time detection of CO and CO_2 for both environmental monitoring and transformer fault detection. Through advanced regression analysis and classification techniques applied to eNose sensor data, the initiative seeks to improve the accuracy and responsiveness of air quality assessments. By addressing the current

limitations of sensor-based technologies, this system presents a scalable and practical alternative to conventional monitoring methods, enabling more effective air quality control and transformer fault diagnostics.

2. Methodology
2.1 Sensor case design for gas sampling

Figure 1 illustrates that the sensor chamber case has been thoughtfully designed to ensure accurate and reliable gas sampling. It features an airtight seal and a controlled gas inlet system, which allows only the intended gas mixture to enter the chamber, thus preventing contamination from outside air and enhancing the precision of the sensor readings. To minimize the risk of leaks, silicone gaskets are used, ensuring that the results remain consistent and accurate. Additionally, the chamber case includes a manual valve, as shown in Figure 2, along with a Tedlar bag, providing precise control over the flow rate and composition of the gas sample, which gives greater confidence in the collected data. Furthermore, the chamber's compact and lightweight design, made possible through 3D printing, makes it cost-effective and easy to use in a variety of settings.
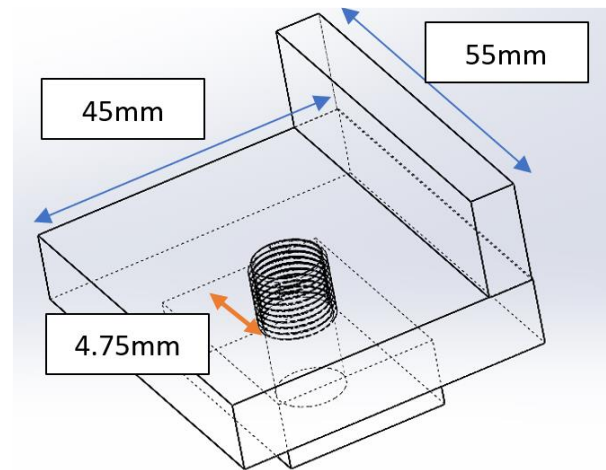
Figure 1. SolidWorks 3D model of the chamber case with central inlet for controlled gas flow
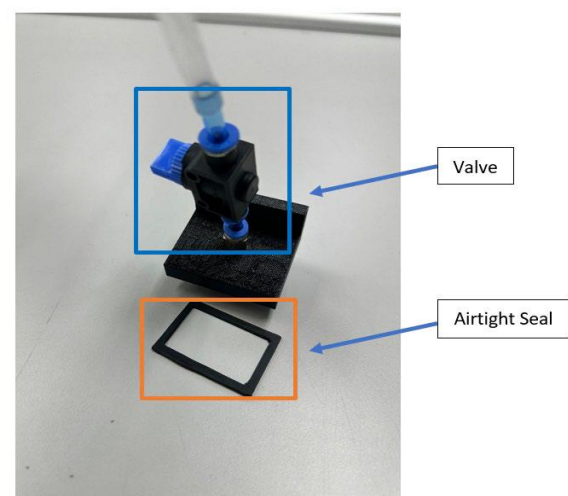
Figure 2. Designed and printed gas sampling chamber case for gas detection

2.2 Digital temperature modulation design by using a buck converter

Figure 3 represents the design process for digital temperature modulation. Before proceeding to full-scale testing and PCB design, each buck converter was first assembled on a breadboard, as shown in Figure 4, to ensure proper functionality. This approach allowed for easy adjustments and troubleshooting within a flexible setup. For the 1A temperature control system, we utilized the LM2575 IC in conjunction with a digital potentiometer to adjust the output voltage, which ranged from 4.0V to 5.0V. The 3A converter, designed for the pneumatic system and based on the LM2576 IC, was tested to confirm a stable 5V output under a 3A load, with an oscilloscope monitoring for any voltage fluctuations. Finally, the 500mA sensor array converter, using the LM2574 IC, provided a consistent 5V output, with careful observation of voltage stability and thermal performance. This prototyping phase confirmed that all converters met the required specifications prior to moving forward with the PCB design.
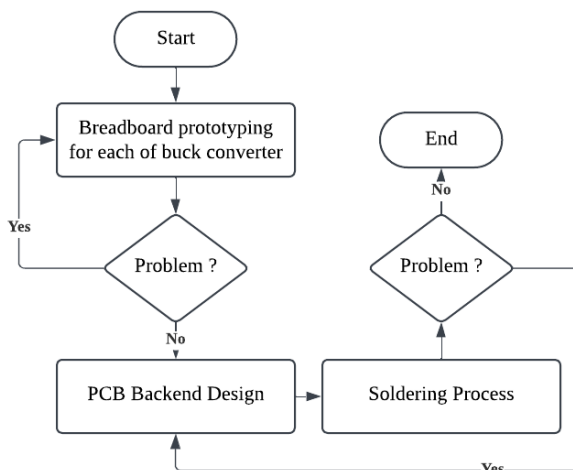

Figure 3. Process of designing the digital temperature modulation

Figure 4. Buck converter breadboard prototyping with the LM2574 500mA IC for troubleshooting

After successfully testing individual buck converters on a breadboard, the next step was to combine all three circuits into a single PCB design. Figure 5 displays the backend PCB layout for combined buck converters. This design aimed to efficiently manage power for the MOS sensor module, pneumatic system, and sensor array. The PCB design process involved converting the schematic into a compact and functional layout. Key considerations included footprint design, component placement, and routing. Accurate pad layouts were created to ensure proper alignment of components and to minimize assembly errors. Components were strategically arranged to reduce wiring complexity while maintaining adequate spacing for airflow and heat dissipation. Routing was done with care, using 1 mm traces for power lines and a copper-filled ground zone to enhance current flow and cooling. The final PCB design measures 10cm x 10cm, a standard size that facilitates easy manufacturing, testing, and future modifications.
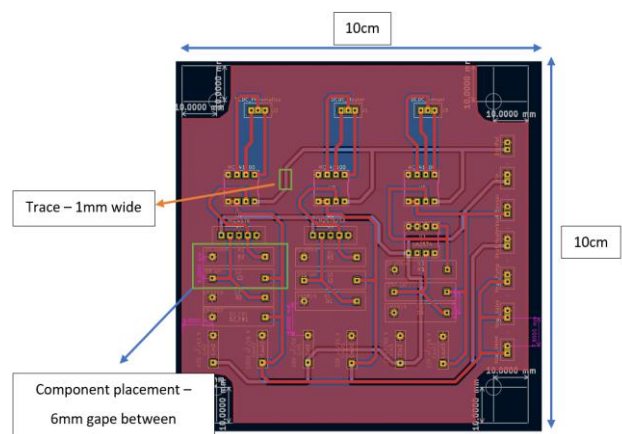

Figure 5. Backend PCB layout for combined buck converters

Once the PCB is prepared, the next step is to solder the components. Figure 6 demonstrates the circuit board with polarized capacitors (highlighted) that require careful orientation during assembly. First, gather the necessary tools: a soldering iron, solder wire, flux, and tweezers. Clean the PCB pads to ensure a smooth soldering process, and then apply flux to help the solder adhere. Carefully place the components in their correct positions, paying special attention to polarized parts like capacitors and diodes.

Start by soldering one pin of each component to hold it in place, and then secure the remaining pins. After soldering, inspect the joints to ensure they are clean and solid. Use a multimeter to test the circuit. Finally, clean off any excess flux and add any finishing touches, such as heatsinks, to complete the process. This ensures everything is securely connected and ready for testing.

2.3 Troubleshooting and resolving inductor overload in the buck converter circuit

Inductors have a maximum current rating, and when this limit is exceeded, they can overheat and fail. In this case, the original inductor was unable to handle the 3A current, leading to its failure. Figure 7 illustrates the replacement process of the 100 µH inductor in the circuit to address the overload

issue. After replacing it with a 100μH inductor rated for 3A, the system functioned correctly again. This experience underscores the importance of selecting components that are properly rated for the system's current demands to ensure reliable performance and prevent damage.
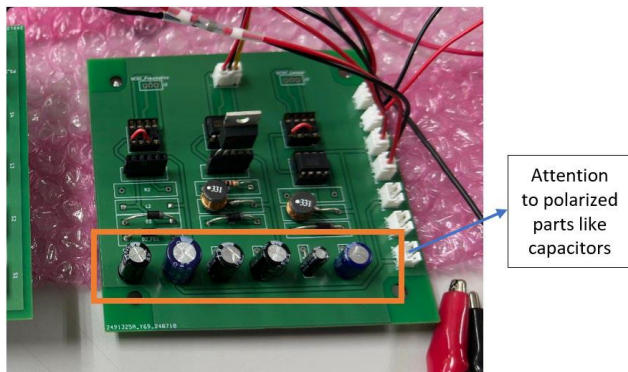
Figure 6. Circuit board with polarized capacitors (highlighted) that require careful orientation during assembly
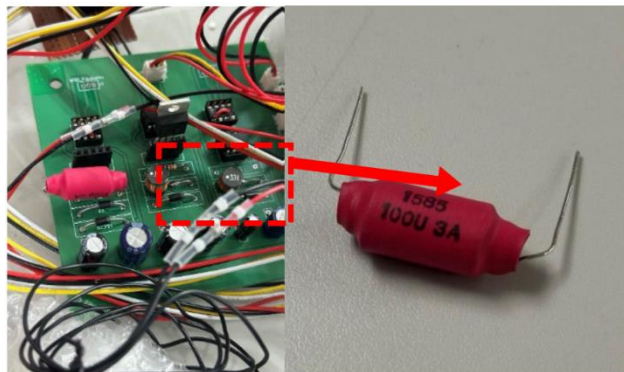
Figure 7. Replace 100μH inductor rated for 3a in the buck converter circuit

2.4 Air sampling process

Flue gas samples were collected using a redesigned Metal Oxide Semiconductor (MOS) gas sensor, a 5V motor pump, and Tedlar bags. Figure 8 illustrates the collection of 75 samples distributed across three air quality categories: good, moderate, and poor. The samples were sourced from clean forested areas, vehicle exhaust points, and gardens with vehicle presence. The motor pump ensured consistent sampling, while the Tedlar bags helped maintain gas purity. Table 1 summarises the distribution of air samples collected across three air quality categories: bad, moderate, and good, with 25 samples gathered for each category, totalling 75 samples. For the bad air quality category, samples were directly collected from the exhausts of four different vehicles: Alza, Bezza, Van, and Myvi, to represent highly polluted environments. The moderate category involved sampling air from four separate areas within the Malaysia-China Friendship Park, a location with a mix of natural and urban influences. For the good category, air was sampled from four locations along the 7th Mile Haji Baki Bamboo Trail, a clean forested area with minimal pollution. All samples were

properly labeled and stored for no more than 24 hours to maintain their integrity before analysis.

Figure 8. Sample collection using Tedlar bags with a total of 75 air samples

Table 1. Air sampling locations and sample distribution

Air Quality Category	Sampling Location	Number of Samples	Store Time
Bad	Directly collect air from 4 different car exhausts (Alza, Bezza, Van, and Myvi).	25	Not over 24 hours
Good	7th Mile Haji Baki Bamboo Trail (Jungle), from 4 areas.	25	
Moderate	Malaysia China Friendship Park, from 4 areas.	25	

Table 2 outlines the timeline and actions involved in the sampling procedure for data collection. Before collecting air samples, the gas sensors are allowed a 30-minute warm-up period to stabilize. During this time, the heater voltage is set to 4.0V, and the system is calibrated to ensure accurate readings. This process is both safe and non-invasive, while the temperature control further enhances sensor sensitivity. The sensors operate with a built-in heater powered by a 5V supply, starting at 4.0V and increasing by 0.1V every 10 seconds until reaching 5.0V at 120 seconds. After that, the voltage drops back to 4.0V, which helps the sensors respond accurately to the air samples. The updated sampling procedure has streamlined the process, reducing the total sample time from 390 seconds to 130 seconds through fully automated voltage adjustments. This improvement enhances efficiency, allowing for more samples to be collected in less time. The 10-step voltage change creates multiple data points,

providing richer and more detailed information, which also improves sensor sensitivity for more accurate detection of pollutant levels.

Table 2. Sampling procedure timeline and actions for data collection

Time	Actions
Connecting Sample	Connect the tube from the Tedlar bag to the chamber. Open the valve from the Tedlar bag first, not the valve from the chamber.
Pre-sampling	The heater automatically sets to 4.0V. Switch the calibration button to "open." The system adjusts the sensors until their output stabilizes at 1.00 ± 0.02V. Once stable, switch the calibration button "off."
0 – 10s	Ensure the output voltage of the sensors is at a steady state at the baseline value, which is around 1.00 ± 0.05V, for 10s.
11s – 20s	Start from 11s, open the chamber valve and gently press the Tedlar bag to let air into the chamber for 10 seconds. Close the chamber valve after 10 seconds.
21s – 120s	The heater voltage automatically steps up to 4.1V at 21 seconds, increasing by 0.1V every 10 seconds until reaching 5.0V at 120 seconds.
121s – 130s	At 121 seconds, the heater voltage automatically steps back down to 4.0V.
Post – sampling	Start the purging process to clear the chamber. Turn on the vacuum valves and air pump until the sensor output stabilizes at 1.00 ± 0.05V. Perform a final check to ensure a steady reading around 1.00 ± 0.02V and get ready for the next sample.

2.5 Data collection

The data collection process involves acquiring raw sensor data from diverse environments with varying air quality conditions. The dataset is systematically divided into two subsets: a training set for model development and a validation set for performance evaluation.

To create the training datasets, a total of 60 air samples were collected and categorized as follows:

- 20 samples of Bad Air,
- 20 samples of Good Air, and
- 20 samples of Moderate Air.

For model validation, an additional 15 air samples were collected, distributed evenly among the categories:

- 5 samples of Bad Air,
- 5 samples of Good Air, and
- 5 samples of Moderate Air.

The electronic nose (eNose) system equipped with Metal Oxide Semiconductor (MOS) sensors was used to capture sensor output voltages over specific time intervals. For regression analysis, a commercial carbon dioxide (CO_2) detector is used to measure and collect CO_2 concentrations, in parts per million (ppm), with five readings at each sampling point.

2.6 Data extraction

Table 3 presents the MOS sensors used in this study, along with their respective targeted gases. Sensor data is recorded continuously from 0 to 130 seconds. However, for analysis, sensor data is extracted from the time intervals between 25 seconds and 115 seconds, in 10-second increments, as shown in Figure 9. This approach selects sensor readings at ten key intervals, capturing critical temporal variations in the sensor's response to different gas concentrations. The extracted data from these intervals is then fed into MATLAB to train and validate the classification models.

Table 3. MOS sensors and targeted gas

Types of MOS sensors	Targeted gas
TGS2600	Hydrogen (H_2), Carbon Monoxide (CO), Methane (CH_4)
TGS2602	VOCs, Ammonia (NH_3), Hydrogen Sulfide (H_2S)
TGS2620	Alcohol, VOCs
TGS2611	Methane (CH_4)
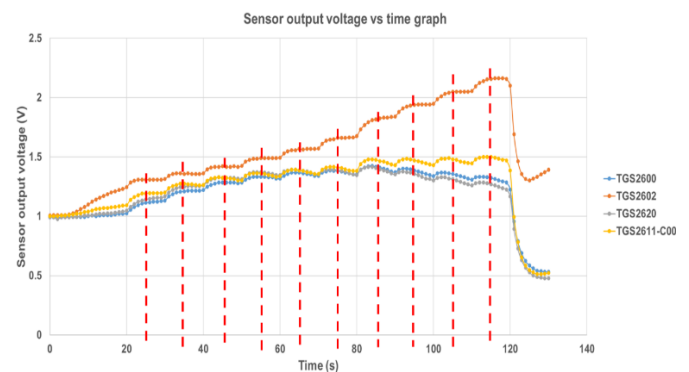
Figure 9. Sensors' instantaneous response extracted at 10-time intervals of the sampling process

2.7 Data analysis

The machine learning classification process, as illustrated in Figure 10 (a), involves a series of steps designed to accurately categorize air quality levels based on sensor data collected from the eNose system. This project will utilize the MATLAB Classification Learner application, which offers 30 classification models. The classification models and their respective model categories are tabulated in Table 4. The process begins with Principal Component Analysis (PCA), which reduces the dimensionality of the raw sensor data while retaining critical features. This step enhances

computational efficiency and improves the overall performance of the classification models. The reduced dataset is utilized to train machine learning models using MATLAB's Classification Learner application, where algorithms such as decision trees, support vector machines, and ensemble methods are assessed to determine the most effective model.

$$Accuracy \ (\%) = \left(\frac{Correct \ predictions}{Total \ predictions} \right) \times 100 \qquad (1)$$

Once the models are trained, data validation is performed using a separate validation dataset to ensure the generalizability and robustness of the models when applied to new, unseen data. The accuracy of each model is then computed using metrics such as training and validation accuracy. Equation (1) is used to compute the accuracy of each model. Following this, the fine-tuning process begins, where hyperparameters are optimized, and the best-performing models are recalibrated to further enhance their performance. Finally, the optimized models are evaluated to confirm their reliability and readiness for deployment in real-world air quality monitoring applications.

(a)   (b)
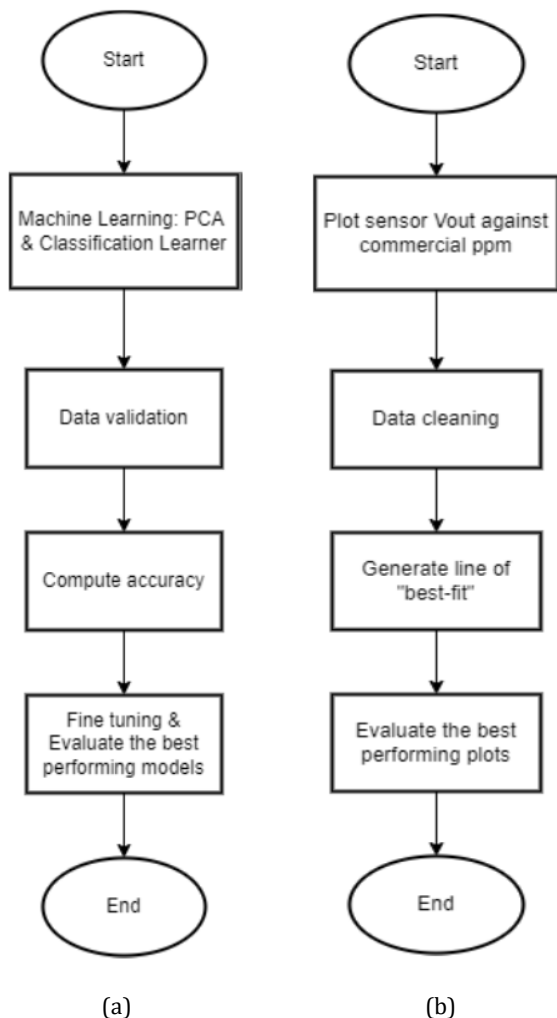
Figure 10. (a) Workflow for the machine learning classification process (b) Workflow for the regression analysis process

The regression analysis workflow, depicted in Figure 10 (b), outlines the steps involved in quantifying pollutant concentrations based on sensor output voltages. The process begins by plotting sensor output voltage (V_{out}) against commercial reference pollutant concentrations (in ppm). Next, the data is cleaned to remove outliers and inconsistencies, ensuring the integrity of the analysis. A line of best fit is then generated using linear regression to model the relationship between V_{out} and pollutant concentrations. Finally, the best-performing regression plots are evaluated for accuracy and reliability, with metrics such as R-squared values used to validate model performance.

Table 4. Classification models and model categories

| Decision Tree	1.Linear SVM	4.Cosine KNN	1.Narrow Neural Network
1. Fine Tree	2.Quadratic SVM	5. Cubic KNN	2.Medium Neural Network
2.Medium Tree	3.Cubic SVM	6.Weighted KNN	3.Wide Neural Network
3. Coarse Tree	4.Fine Gaussian SVM	**Ensemble Classifiers**	4.Bilayered Neural Network
Discriminant Analysis	5.Medium Gaussian SVM	1.Boosted Trees	5.Trilayered Neural Network
1.Linear Discriminant	6.Coarse Gaussian SVM	2.Bagged Trees	**Kernel Approximation Classifiers**
Naïve Bayes Classifier	**Nearest Neighbour Classifiers**	3.Subspace Discriminant	1. SVM Kernel
1.Gaussian Naïve Bayes	1. Fine KNN	4.Subspace KNN	2.Logistic Regression Kernel
2.Kernel Naïve Bayes	2.Medium KNN	5.RUS Boosted Trees	
Support Vector Machines (SVM)	3.Coarse KNN	**Neural Network Classifiers**	

3.   Results and discussion

3.1   Digital temperature modulation output voltage testing

After integrating the buck converters onto the PCB, voltage regulation testing was conducted to verify their output voltages. The voltage regulation test of LM2575 Buck converter on PCB was setup as shown in Figure 11. LM2576 provided a stable 5.07V, suitable for high-current applications, while LM2574 delivered 5.05V for low-current applications. The LM2575, controlled by an Arduino, adjusted its output from 4.0V to 5.0V in 0.1V increments,

demonstrating precise voltage control for temperature modulation. Table 5 summarises the output voltage testing results of the LM2576, LM2574, and Arduino-controlled LM2575 buck converters. These results confirm the successful integration and reliable performance of all three converters.
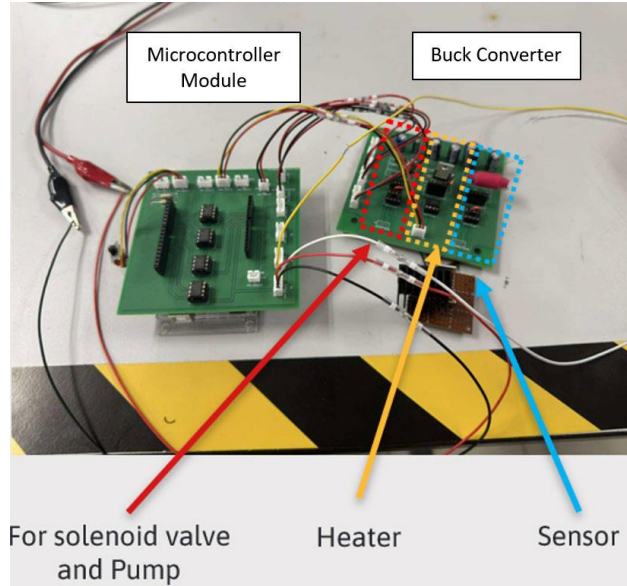

Figure 11. Voltage regulation test of the LM2575 buck converter on the PCB setup

Table 5. Output voltage testing results of LM2576, LM2574, and arduino controlled LM2575 buck converters

Buck Converter IC	Input Voltage (V)	Time (s)	Output Voltage (V)
3A Current LM2576	12	-	5.07
500mA Current LM2574	12	-	5.05
1A Current LM2575	12	0-19	4.02
		20-29	4.11
		30-39	4.21
		40-49	4.31
		50-59	4.41
		60-69	4.50
		70-79	4.60
		80-89	4.74
		90-99	4.82
		100-109	4.94
		110-119	5.03
		120-129	4.02

3.2 Evaluating the digital temperature modulation sensor response in a graph

This test evaluated the performance of the Digital Temperature Control system, which is designed to improve sensor accuracy, consistency, and efficiency. The performance of this system is presented in Figure 12. During the test, the heater voltage was gradually increased by 0.1V every 10 seconds, starting at 4.0V and reaching a peak of 5.0V, before smoothly returning to 4.0V. The response curve indicated that the sensors (TGS2600, TGS2602, TGS2620, and TGS2611) produced stable and consistent outputs throughout the process. The updated process improved sampling consistency and reduced sampling time. While the previous design took 390 seconds to complete each sample, the automated system now accomplishes this in just 130 seconds. This threefold reduction not only saves time but also allows researchers to collect more samples, ultimately increasing overall efficiency and productivity. The gradual voltage adjustments ensured that the sensors responded steadily, resulting in accurate and uninterrupted data collection. Furthermore, the system's capability to adjust voltage without delays or fluctuations underscores its reliability. These results demonstrate that the Automatic Temperature Control system effectively enhances sampling performance while making the process faster and more efficient.
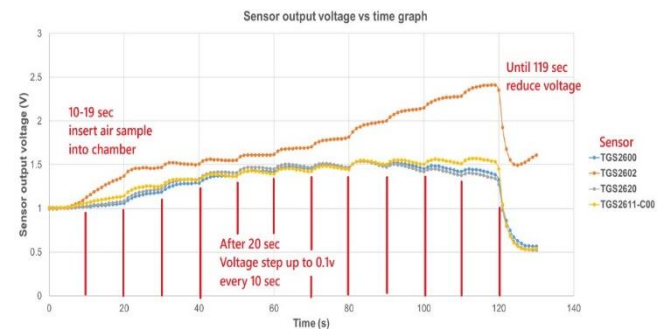

Figure 12. Performance of the digital temperature control system

3.3 Sensor performance and stability in different air quality

The response curves offer valuable insights into how the Metal Oxide Semiconductor (MOS) gas sensors react to different air quality conditions, highlighting their effectiveness in identifying and classifying pollutant levels. Figures 13 and Figure 14 show the response curves for bad air quality samples, where sensors (TGS2600, TGS2620, and TGS2611) react strongly to high concentrations of CO_2 and CO. The voltage rises rapidly, reaching between 3V and 5V within the first 10 seconds after flue gas is introduced into the chamber. TGS2600 and TGS2620 maintain steady performance throughout the sampling period. However, TGS2611 shows less reliable behavior after 20 seconds, with its voltage dropping to a range of 2.5V to 4V by 120 seconds. This decline is likely due to changes in air saturation and temperature variations, which impact pollutant levels and sensor accuracy.
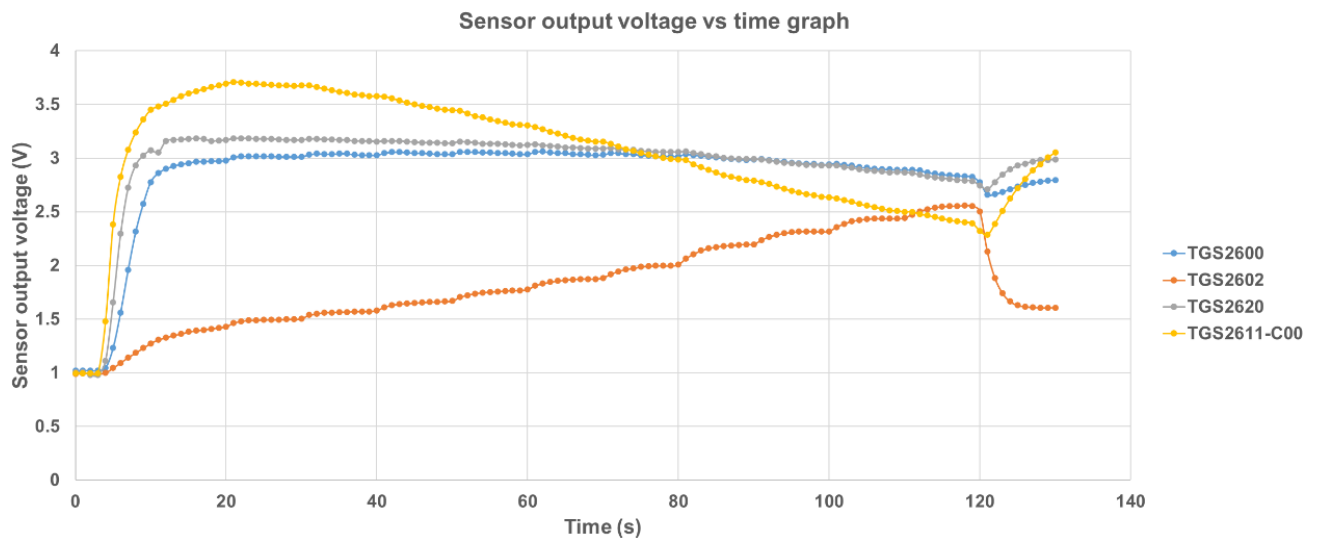
Figure 13. Shape of sensor response curve showing instability and voltage drop over time in bad air quality (sample 2)
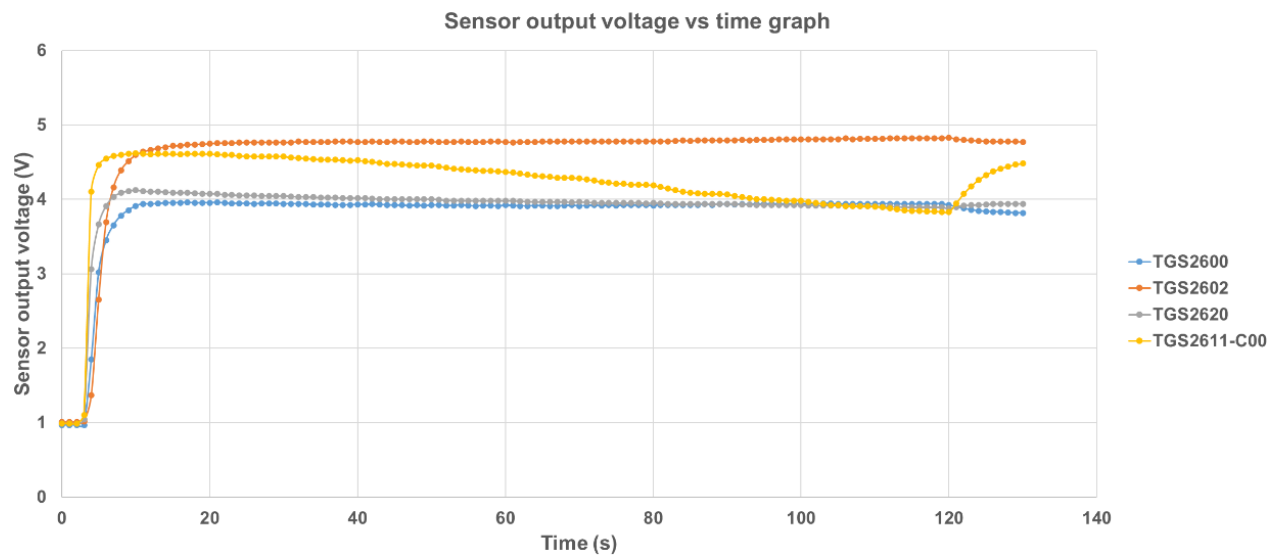
Figure 14. Shape of sensor response curve showing instability and voltage drop over time in bad air quality (sample 8)

For good air quality samples, as shown in Figure 15, the sensors maintain stable voltage levels below 2.5V, reflecting low concentrations of CO_2 and CO in the clean air. This steady performance indicates the sensors' reliability in detecting minimal pollutant levels. In Figure 16, moderate air quality samples also exhibit steady voltage responses, like good air quality. The voltages remain below 2.5V, placing all moderate samples in the clean air category. A clear distinction emerges when comparing bad air samples to good and moderate ones. Bad air samples produce significantly higher voltage responses due to the higher concentrations of CO_2 and CO, while good and moderate samples yield much lower and more stable outputs.

These findings confirm the sensors' ability to reliably differentiate between varying air quality conditions, providing consistent and accurate responses for classifying pollutants.

3.4 Sensor accuracy analysis

Table 6 presents the accuracy of four gas sensors: TGS2600, TGS2602, TGS2620, and TGS2611, in detecting gases at various time intervals ranging from 25 to 115 seconds. The results indicate that optimal performance occurred between 45 and 65 seconds, where all sensors achieved their highest accuracy levels. TGS2602 recorded the highest accuracy of 65.33% at both 45 and 65 seconds, followed closely by TGS2600, which peaked at 62.67% at 55 and 65 seconds.

39
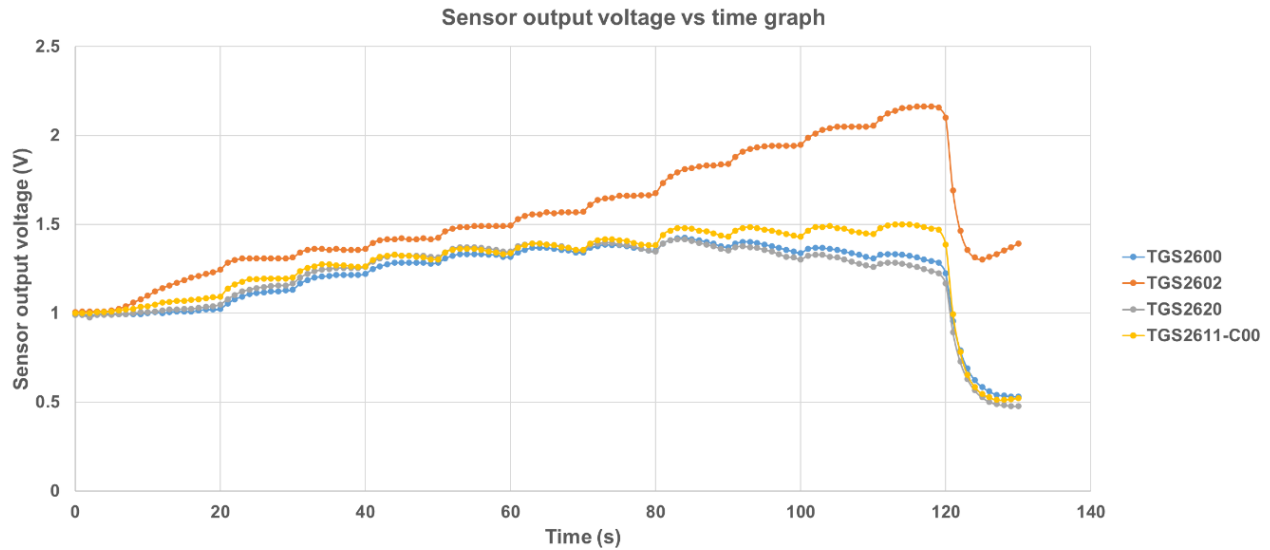
Figure 15. Shape of sensor response curve showing stable voltage over time in good air quality (sample 24)
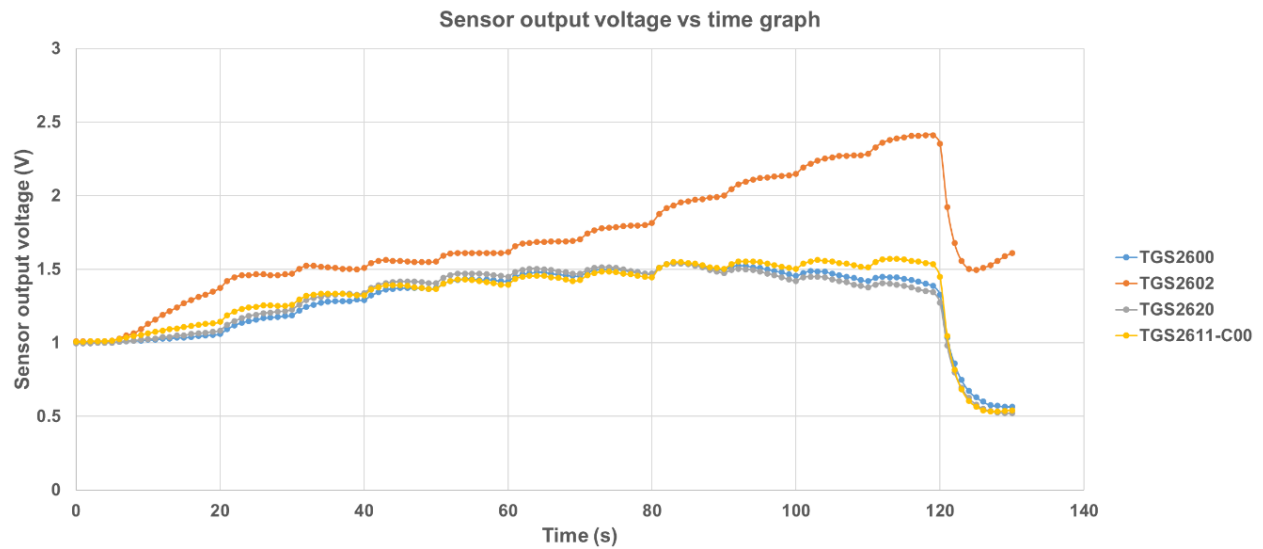
Figure 16. Shape of sensor response curve showing stable voltage over time in moderate air quality (sample 56)

TGS2620 reached its maximum of 60.00% at 65 seconds, while TGS2611 performed best at 45 seconds with an accuracy of 58.67%. After 75 seconds, the accuracy of most sensors either declined or stabilized, with TGS2602 notably dropping to a consistent 52.00% from 85 seconds onward. These findings suggest that the 45- to 65-second window is the most effective period for capturing accurate gas detection data, underscoring the importance of timing in optimizing sensor performance. Table 7 presents the refined analysis of gas sensor accuracy for four sensor models: TGS2600, TGS2602, TGS2620, and TGS2611, measured across time intervals from 25 to 115 seconds. The results show a notable improvement in detection performance compared to the initial analysis. TGS2600 maintained a high and consistent accuracy of 92.86% from 25 to 55 seconds before stabilizing at 91.43% for the remainder of the test.

TGS2602 improved to 91.43% from 45 seconds onward, while TGS2620 demonstrated consistent accuracy, peaking at 92.86% at 25 seconds and maintaining 91.43% throughout most subsequent intervals. TGS2611 maintained an accuracy of 91.43% across nearly all time points, with a slight increase to 92.86% at 105 seconds. Overall, all four sensors exhibited strong and stable performance with minimal fluctuation after 45 seconds, indicating improved system stability and sensor reliability under refined testing conditions. At 25 seconds, TGS2600 and TGS2620 recorded the highest accuracy of 92.86%, effectively detecting and classifying air quality as good or bad. TGS2602 and TGS2611 followed closely with accuracies of 90.00% and 91.43%, respectively. Over time, all sensors stabilized at 91.43%, demonstrating reliable and consistent detection capability. These findings confirm the sensors' ability to accurately differentiate between air quality conditions. TGS2600 and TGS2620 showed superior early

detection performance, making them particularly well-suited for both short-term and long-term air quality monitoring in real-world applications.

Table 6. The accuracy of the gas sensors in correctly detecting gases (initial analysis)

Time (s)	TGS2600	TGS2602	TGS2620	TGS2611
	Accuracy (%)			
25	53.33	60.00	50.67	52.00
35	50.67	61.33	48.00	49.33
45	54.67	65.33	54.67	58.67
55	62.67	64.00	57.33	56.00
65	62.67	65.33	60.00	52.00
75	58.67	53.33	58.67	49.33
85	56.00	53.33	54.67	52.00
95	54.67	52.00	54.67	54.67
105	57.33	52.00	56.00	54.67
115	61.33	52.00	56.00	52.00

Table 7. The accuracy of gas sensors in correctly detecting gases (refined analysis)

Time (s)	TGS2600	TGS2602	TGS2620	TGS2611
	Accuracy (%)			
25	92.86	90.00	92.86	91.43
35	92.86	90.00	91.43	91.43
45	92.86	91.43	91.43	91.43
55	92.86	91.43	91.43	91.43
65	91.43	91.43	91.43	91.43
75	91.43	91.43	91.43	91.43
85	91.43	91.43	91.43	91.43
95	91.43	91.43	91.43	91.43
105	91.43	91.43	91.43	92.86
115	91.43	91.43	90.00	91.43

The sensor response to different air quality samples categorized as "bad," "good," and "moderate" is illustrated in Figure 13, Figure 14, Figure 15, and Figure 16, respectively. These graphs correspond to Sample 2, Sample 8, Sample 24, and Sample 56. Each graph plots the sensor output voltage over time (in seconds) for the four Metal Oxide Semiconductor (MOS) sensors used in the study. For the bad air sample (Sample 2 & 8), the sensor output voltage rises sharply and stabilizes at high levels, reflecting strong detection of high pollutant concentrations typical of poor air quality. For the good air sample (Sample 24), voltage readings are significantly lower, stabilizing at low levels, indicating reduced pollutants and cleaner air. For the moderate air sample (Sample 56), voltage readings fall between the bad and good samples, showing a moderate rise and stabilization, effectively capturing intermediate air quality conditions. These results confirm the sensors' ability to differentiate between varying pollution levels.

3.5 Classification results

The best-performing classification models and their corresponding validation and classification accuracies are summarized in Table 8. The table highlights the top-performing classification models: Bagged Trees, Tri-layer Neural Network, SVM Kernel, and Logistic Regression Kernel. All models achieved a classification accuracy of 100%, correctly classifying all training data without misclassification. For validation accuracy, each model recorded 93.33%, indicating strong generalization and consistent reliability when applied to unseen data. The confusion matrices for the four best-performing models (Bagged Trees, Tri-layer Neural Network, SVM Kernel, and Logistic Regression Kernel) are shown in Table 9. These matrices provide a detailed view of each model's ability to classify air quality samples into "Bad," "Good," and "Moderate" categories. Among the models, Bagged Trees demonstrated the best performance, with minimal misclassifications. Specifically, only one "Good" sample was classified as "Moderate," and one "Moderate" sample was classified as "Good," reflecting high reliability and robustness. The Tri-layer Neural Network also achieved strong results, but it exhibited slightly more errors, misclassifying three "Bad" samples as "Good" and one "Moderate" sample as "Good," suggesting reduced sensitivity in identifying poor air quality. The SVM Kernel model performed consistently, with balanced classification across categories. However, it struggled with borderline cases, misclassifying two "Bad" samples as "Moderate" and one "Good" sample as "Moderate." Logistic Regression Kernel showed the highest misclassification rates, with two "Bad" samples classified as "Good" and one "Moderate" sample classified as "Good," indicating reduced robustness when differentiating between similar air quality categories. Despite all models achieving 100% classification accuracy during training, their validation accuracy – 93.33% for each model – reveals differences in their ability to generalize to unseen data. Bagged Trees stood out as the most reliable model for distinguishing air quality levels, followed closely by the Tri-layer Neural Network and SVM Kernel. Logistic Regression Kernel, while effective, exhibited more challenges in handling nuanced distinctions between air quality categories. These results confirm Bagged Trees as the most effective and robust model for this classification task.

Table 8. Best performing classification models

Types of Classification Models	Best Performing Models	
	Validation accuracy (%)	Classification accuracy (%)
Bagged Trees	93.33	100
Trilayer Neural Network	93.33	100
SVM Kernel	93.33	100
Logistic Regression Kernel	93.33	100

Table 9. Confusion matrices of best-performing models

Confusion Matrix	
Bagged Trees	**SVM Kernel**

Tri-layer Neural Network	**Logistic Regression Kernel**
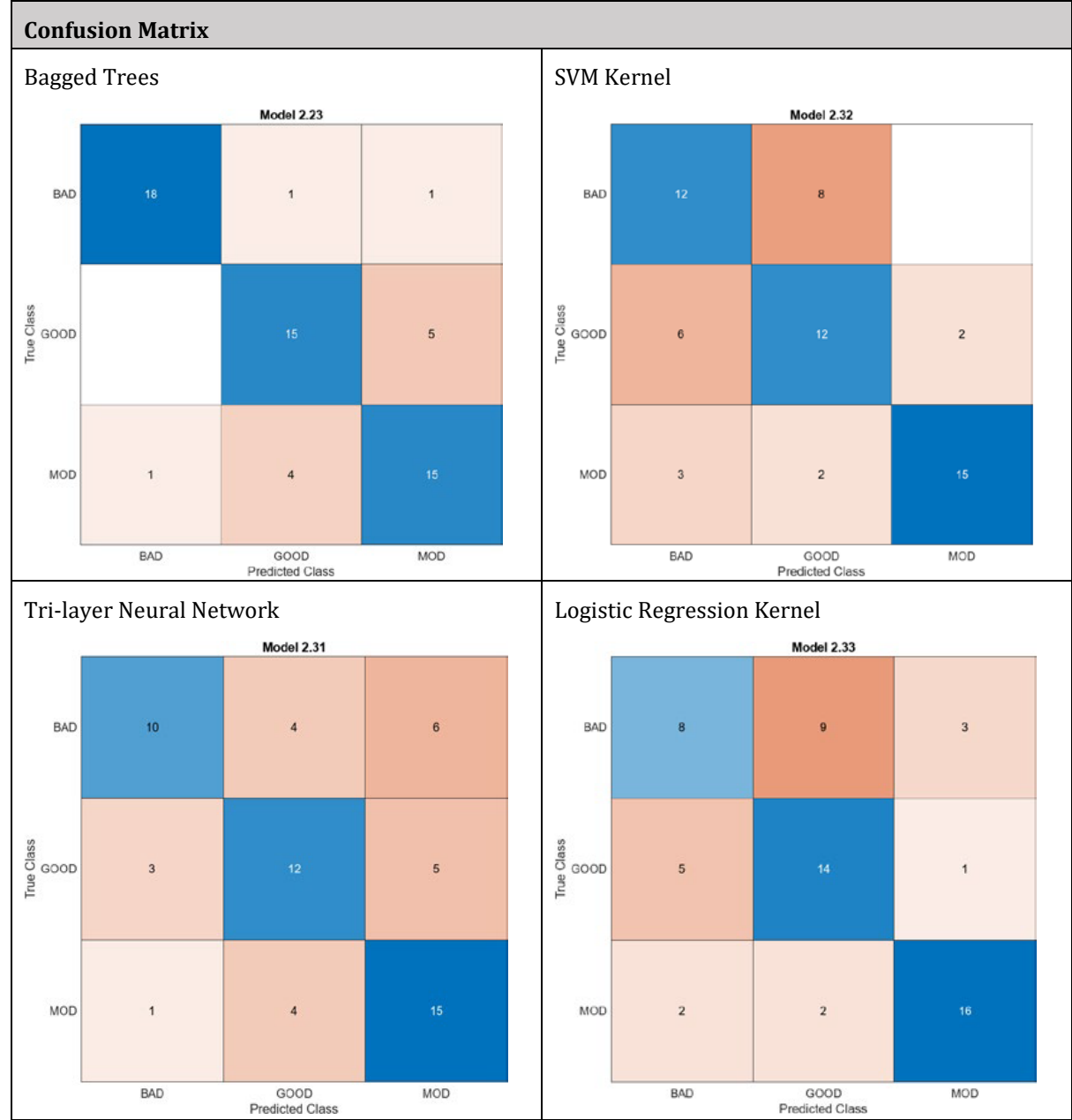
Table 1. CO_2 Concentrations at Kampung Haji Baki

	Collection Points			
	Kampung A	**Kampung B**	**Kampung C**	**Kampung D**
CO_2 concentrations (ppm)	4525	3735	3100	1930
	4045	3670	3155	1885
	4430	3880	1880	1745
	4310	4115	1835	1580
	4270	3845	1855	1359
Average (ppm)	4316	3849	2365	1699.8

3.6 Linear regression results

For air samples with CO_2 concentrations exceeding the 5000 ppm detection limit of standard sensors (e.g., car exhaust), measurements were capped at 5000 ppm for regression modelling. This simplifies analysis by standardizing maximum values while acknowledging the sensor's limitations. Although variability above 5000 ppm is not captured, this approach ensures meaningful insights without introducing bias from non-detectable values. The regression analysis was performed to quantify CO_2 concentration levels (in ppm) across different collection points using sensor output voltages. The analysis focused on two key locations: Kampung Haji Baki and the Malaysia-China Friendship Park, with the results summarized in Table 10 and Table 11, respectively. Table 10 shows CO_2 concentrations at Kampung Haji Baki, categorized as "good air," with averages ranging from 1,699.8 ppm to 4,316 ppm, reflecting low pollutant levels. Table 11 presents CO_2 concentrations at Malaysia-China Friendship Park, categorized as "moderate air," with averages between 459 ppm and 565 ppm, indicating slightly higher pollutant levels.

The sensor output voltage was plotted against commercial PPM values across 10 time intervals (25s to 115s). The R-squared values, indicating the goodness of fit for the linear trendline, were calculated for each sensor (TGS2600, TGS2602, TGS2620, TGS2611). Higher R-squared values (closer to 1) signify better consistency and reliability of the data. Outliers were removed by identifying significant deviations from mean values and overlapping points to prevent skewed regression analysis. This process improved regression accuracy, increased R-squared values, and enhanced the predictive reliability of the models. As shown in Table 12, outlier removal significantly improved the R-squared value, indicating better consistency in the data. For further analysis, three intervals – 25s, 45s, and 75s – were selected due to their relatively high R-squared values. These intervals demonstrated strong linear relationships across all sensors, ensuring robust and reliable trends for interpreting results and discussion. TGS2602 sensor consistently achieved the highest R-squared values across the selected time intervals (25s, 45s, 75s), as shown in Table 13.

Table 11. CO_2 concentrations at Malaysia–China Friendship Park

	Collection Points			
	China A	**China B**	**China C**	**China D**
CO_2 concentrations (ppm)	440	595	410	470
	430	575	410	505
	420	550	680	525
	455	555	465	545
	550	550	415	540
Average (ppm)	459	565	476	517

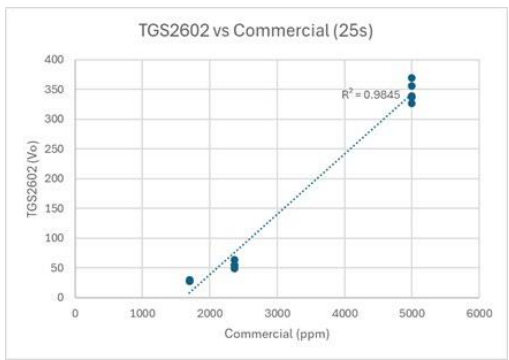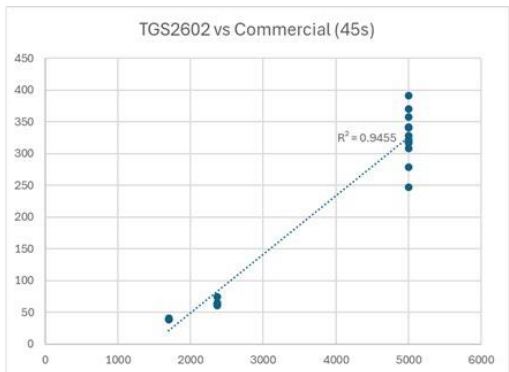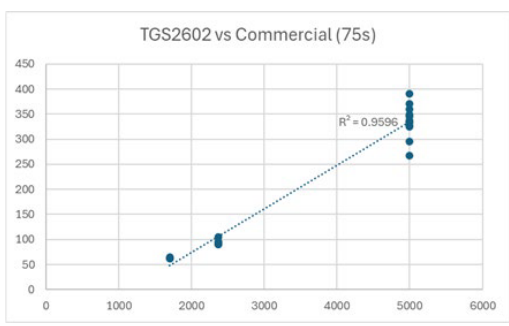**Table 12.** Comparison of V_{out} vs. commercial (ppm) at 25-th second interval for TGS2602 sensor

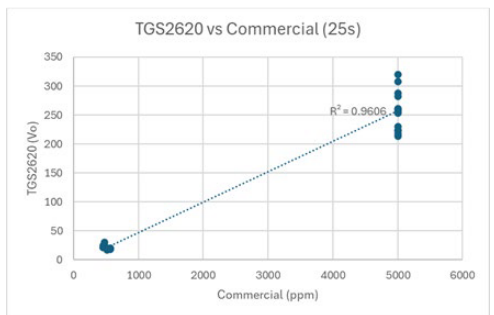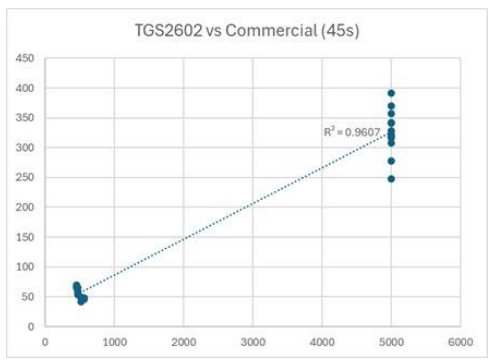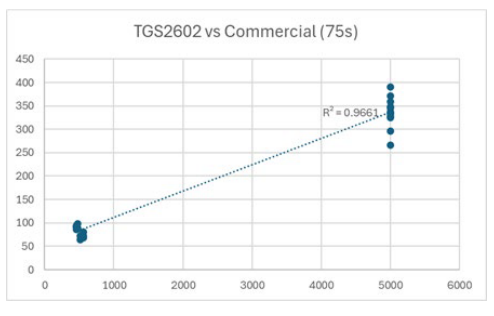| **Before Outlier Removal:**	**After Outlier Removal:**
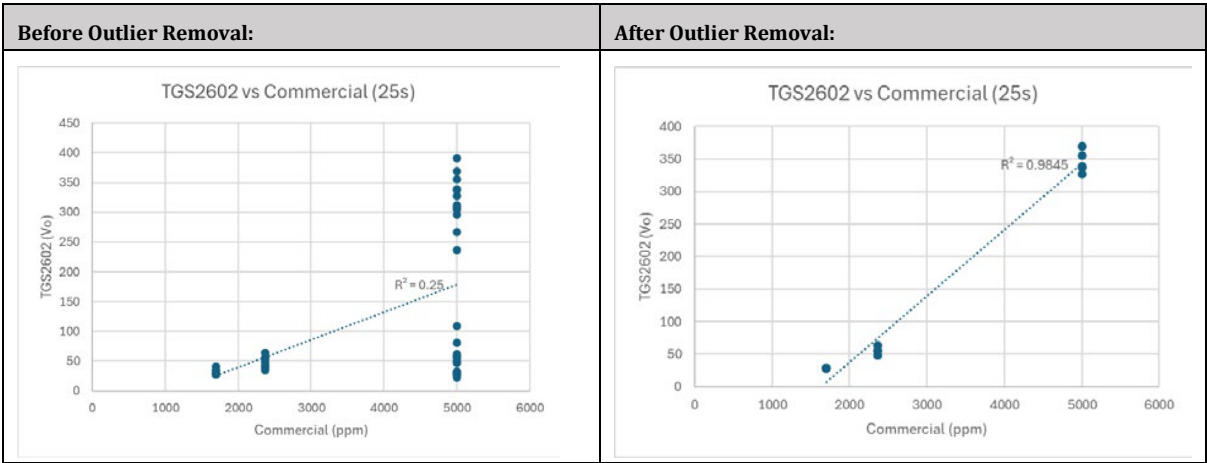
These intervals exhibit the strongest linear relationships between V_{out} and commercial PPM values, establishing TGS2602 as the most reliable sensor for regression analysis in this study. Furthermore, while commercial CO_2 sensors are limited to a maximum concentration of 5000 ppm, the MOS gas sensors used in this study demonstrate the capability to measure beyond this range, highlighting their extended detection potential. For moderate and bad samples, TGS2602 and TGS2620 recorded the highest R-squared values among the four sensors at the selected time intervals (25s, 45s, 75s), as shown in Table 14. To highlight the most reliable data, this paper presents the Vout vs. PPM plots for TGS2620 at the 25s interval and TGS2602 at the 45s and 75s intervals. These intervals demonstrate the strongest linear relationships, making them ideal for regression analysis.

Referring to Table 10 and Table 11, CO_2 concentrations differ between "good air" and "moderate air" quality levels. At Kampung Haji Baki (Table 10), average CO_2 concentrations range from 1699.8 ppm to 4316 ppm, categorized as good air. In contrast, Malaysia-China Friendship Park (Table 11) shows averages from 459 ppm to 565 ppm, reflecting moderate air quality. These variations highlight the impact of location on CO_2 levels. Forests like Kampung Haji Baki have higher CO_2 due to soil respiration, microbial activity, and restricted airflow from dense vegetation. Urban parks, such as Malaysia-China Friendship Park, benefit from photosynthesis, better airflow, and aquatic CO_2 absorption, resulting in lower CO_2 levels (400–700 ppm).

Table 13. Plot of V_{out} vs. commercial (ppm) for good and bad samples at 25s, 45s and 75s

Time interval	Plot of V_{out} vs Commercial (ppm)
25s	
45s	
75s	

Table 14. Plot of V_{out} vs. commercial (ppm) for moderate and bad samples at 25s, 45s and 75s

Time interval	Plot of V_{out} vs Commercial (ppm)
25s	
45s	
75s	

4. Conclusions

This research demonstrated the development and effectiveness of an advanced eNose system, equipped with Metal Oxide Semiconductor (MOS) sensors, for real-time air quality and flue gas monitoring. By addressing key limitations of traditional gas sensing systems, the study introduced a more efficient, accurate, and cost-effective solution for detecting pollutants such as carbon monoxide (CO) and carbon dioxide (CO_2). Key advancements included precise temperature control, improved sensor chamber design, and optimized sampling methods, all of which enhanced the performance and stability of the MOS sensors. The integration of automatic digital buck converters significantly reduced the sampling time from 390 seconds to just 130 seconds, while maintaining consistent and reliable operation. The redesigned sensor chamber ensured improved airflow, maintained airtight conditions, and minimized external interference, resulting in more accurate gas detection. The system successfully quantified pollutant concentrations and classified air quality into "good," "moderate," and "bad" categories with over 91% classification accuracy. Machine learning models, such as Bagged Trees and SVM Kernel, further improved classification reliability, while linear regression models demonstrated strong correlations, confirming the accuracy of sensor outputs. Analysis of sensor behavior revealed that poor air quality was characterized by rapid voltage increases (3V–5V), while good to moderate air conditions showed stable voltages below 2.5V. Notably, 45-second sampling intervals provided an optimal balance between sensor response stability and classification accuracy. Additionally, outlier removal significantly improved data quality and model performance. Effective strategies, such as the use of Tedlar bags for sample preservation and automated voltage adjustments, enhanced the consistency of air sample collection. Looking ahead, future research should explore the eNose's performance under varied environmental conditions, including fluctuations in temperature and humidity, to ensure greater robustness. Refining sampling methodologies for moderate and high-pollution scenarios, incorporating advanced sensors with extended detection ranges, and conducting long-term field deployments will further align the system with conventional air monitoring standards. In conclusion, the eNose system represents a scalable, portable, and reliable tool for environmental monitoring and industrial fault detection. Its high efficiency, adaptability, and real-time capabilities make it a valuable asset for modern air pollution management strategies.

Ethical issue
The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement
The manuscript contains all the data. However, more data will be available upon request from the corresponding author.

Conflict of interest
The authors declare no potential conflict of interest.

References
[1] Rabia, R., et al., "Impact of industrial processes on air quality and public health," Journal of Environmental Science, vol. 29, no. 4, pp. 215–223, 2021.

[2] H. Ritchie. "Deaths from air pollution are high, but the data contains hope." Clean air fund. https://www.cleanairfund.org/news-item/deaths-air-pollution-data-hope/#:~:text=The%20World%20Health%20Organization%20estimates,from%20burning%20wood%20and%20charcoal (accessed 29 November 2024, 2024).

[3] H. E. Lee, H. S. Chua, Z. J. A. Mercer, S. M. Ng, and M. Shafiei, "Fraud detection of black pepper using metal oxide semiconductor gas sensors," in 2021 IEEE Sensors, 31 Oct.-3 Nov. 2021, pp. 1-4, doi: 10.1109/SENSORS47087.2021.9639658.

[4] H. E. Lee, Z. J. A. Mercer, S. M. Ng, M. Shafiei, and H. S. Chua, "Metal Oxide Semiconductor Gas Sensors-based E-nose and Two-stage Classification: Authentication of Malaysia and Vietnam Black Pepper Samples," in 2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), 29 May-1 June 2022 2022, pp. 1-4, doi: 10.1109/ISOEN54820.2022.9789618.

[5] Hui En Lee et al., "Temperature modulation of metal oxide semiconductor gas sensors and machine learning for geo-tracing of food products and classification of transformer oil quality," vol. 1, Exploring engineering: an anthology of multidisciplinary undergraduate research, Colin Choon Lin Tan and B. T. Lau, Eds., Sarawak, Malaysia: Swinburne Sarawak Sdn. Bhd., 2023, pp. 104 - 119. [Online]. Available: https://swinburne.librarynet.com.my/Angka.sa2/swinburne/OpacBibDetail.htm?bibId=742488

[6] A. Raihan, R. A. Begum, M. N. Mohd Said, and J. J. Pereira, "Assessment of Carbon Stock in Forest Biomass and Emission Reduction Potential in Malaysia," Forests, vol. 12, no. 10, p. 1294, 2021. [Online]. Available: https://www.mdpi.com/1999-4907/12/10/1294.

[7] A. Cavaliere et al., "Development of Low-Cost Air Quality Stations for Next Generation Monitoring Networks: Calibration and Validation of PM2.5 and PM10 Sensors," Sensors, vol. 18, no. 9, p. 2843, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/9/2843.

[8] Sun, W., et al., "Discriminative detection of different cigarette brands using a fast-response electronic nose," ACS Omega, vol. 8, pp. 46034–46042, 2023.

[9]   Karim, S 2021, 'Train and evaluate a classification model in machine learning!', Medium, viewed 12 June 2024, < https://medium.com/@sarakarim/train-and-evaluate-a-classification-model-in-machine-learning-18fbd6504da3>.

[10]   H. E. Lee, Z. J. A. Mercer, S. M. Ng, M. Shafiei, and H. S. Chua, "Metal Oxide Semiconductor Gas Sensors-based E-nose and Two-stage Classification: Authentication of Malaysia and Vietnam Black Pepper Samples," in 2022 IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), 29 May-1 June 2022 2022, pp. 1-4, doi: 10.1109/ISOEN54820.2022.9789618. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9789618

[11]   A. Sudarmaji and A. Kitagawa, "Application of Temperature Modulation-SDP on MOS Gas Sensors: Capturing Soil Gaseous Profile for Discrimination of Soil under Different Nutrient Addition," (in English), Journal of Sensors, Article p. NA, 2016 Annual // 2016. [Online]. Available: https://link.gale.com/apps/doc/A513009155/AONE?u=anon~e45f8757&sid=googleScholar&xid=90bef4d5