A Prototype Sensor-Based System with Machine Learning for Cannabis Strain Classification via VOC Signatures

Suparat Sasrimuang[1], Thanyathorn Ninduangdee[2], Teerapat Khottarin[2], Suchin Trirongjitmoah[2] and Nisarut Phansiri[2, *]
[1] Chemical Engineering Department, Engineering Faculty, Ubon Ratchathani University, Si Khai, Warin Chamrap, Ubon Ratchathani, 34190, Thailand
[2] Electrical and Electronics Department, Engineering Faculty, Ubon Ratchathani University, Si Khai, Warin Chamrap, Ubon Ratchathani, 34190, Thailand
***Corresponding Author E-mail:** nisarut.p@ubu.ac.th
Received: Mar 30, 2025; **Revised:** Jun 04, 2025; **Accepted:** Jun 11, 2025

Abstract

Cannabis strain classification is essential for ensuring product consistency, therapeutic accuracy, and quality control in both medical and commercial applications. This study presents the development of a prototype sensor-based system for classifying three cannabis strains—Black Patronas (Hybrid), MAC Gold (Indica), and Banana Daddy R1 (Sativa)—by analyzing odors emitted from dried flowers and leaves. A gas sensor array consisting of five low-cost sensors (MQ-2, MQ-3, MQ-6, TGS-822, and TGS-826) was employed to detect volatile organic compounds (VOCs) characteristic of each strain. Sensor signals were acquired using an Arduino Mega 2560, preprocessed via Node-RED, stored in InfluxDB, and visualized using Grafana. To enable classification, differential responses (ΔR) were computed by subtracting baseline analog values from VOC exposures. These ΔR values were used to train a Random Forest classifier, which achieved an accuracy of 83% on unseen test samples. Notably, MQ-2 showed strong response to Hybrid flowers, while TGS-822 was most effective for detecting VOCs from Sativa leaves. While the dataset was limited in size, the system demonstrated reliable classification across six cannabis sample types. These results confirm the feasibility of using low-cost sensor arrays with interpretable ML models for odor-based strain identification. Future work will expand the dataset, explore additional algorithms, and move toward real-time, portable deployment for cannabis quality assurance.

Keywords: Cannabis strain classification, Volatile organic compounds (VOCs), Gas sensor array, Electronic nose, Machine learning

1. Introduction

The global surge in cannabis sativa legalization has transformed the plant from a stigmatized substance into a billion-dollar agricultural commodity [1]. Cannabis is now central to medical therapies, wellness products, and industrial applications. With this rapid growth comes an urgent demand for precise strain classification to ensure product safety, therapeutic efficacy, and regulatory compliance [2],[3]. Each cannabis strain-Indica, Sativa, and Hybrid—contains distinct profiles of cannabinoids, terpenes, and volatile organic compounds (VOCs) that directly influence physiological effects. Mislabeling or misidentification of strains not only undermines consumer trust but can also compromise medical outcomes [4–7].

Despite its importance, current strain identification methods are heavily reliant on sophisticated laboratory techniques such as chromatography and genetic testing. These methods, while accurate, are costly, time-intensive, and unsuitable for on-site or real-time deployment—especially in cultivation fields, dispensaries, or border control environments [8],[9]. This technological gap poses a barrier to both quality control and legal enforcement.

In the context of Thailand, cannabis was officially decriminalized in 2022, marking a significant shift in national drug policy. The removal of cannabis from the list of controlled narcotics has opened new opportunities for its medical, commercial, and research applications. However, this rapid liberalization also raises concerns regarding regulation, public health, and product quality control. With a growing number of cannabis cultivators and products entering the Thai market, there is an increasing need for reliable and cost-effective tools to authenticate and classify cannabis strains. Strain misidentification may lead to misuse, compromised therapeutic outcomes, or legal disputes, especially when THC content is a regulatory factor. Thus, strain classification is not only a technical issue but also a public and legal necessity.

Recent advances in gas sensor arrays—often referred to as electronic noses—have opened new pathways for odor-based classification [10–12]. When integrated with machine learning (ML), these systems can detect subtle patterns in VOC emissions, enabling automated and intelligent decision-making. Such platforms promise a portable, affordable, and scalable solution to a long-standing challenge in the cannabis industry [13–15].

This study presents the design and implementation of a prototype sensor-based system for cannabis strain

classification. By combining a multi-sensor gas array with embedded data processing (Arduino Mega 2560), real-time visualization (Grafana), and ML-based prediction, the system aims to distinguish among three commercial strains: Black Patronas (Hybrid), MAC Gold (Indica), and Banana Daddy R1 (Sativa). The analysis includes both dried flowers and leaves, making the approach comprehensive and practical. This work contributes to the foundation of smart, field-deployable cannabis authentication systems—marking a step forward in agricultural sensing, AI-based classification, and cannabis quality assurance.

2. Theoretical Background

2.1 Principles of Gas Sensors (MQ and TGS Series)

Gas sensors detect specific types of gaseous molecules by converting chemical interactions into electrical signals. The MQ and TGS series sensors operate based on changes in the resistance of a sensing material when exposed to gases. MQ sensors typically use a SnO_2 (tin dioxide) semiconductor, which changes its electrical resistance in the presence of gases such as alcohol, methane, or hydrogen. TGS (Taguchi Gas Sensors), developed by Figaro, employ similar principles but are tuned to different gas sensitivities. For example, TGS-822 is particularly responsive to organic solvents, while TGS-826 targets ammonia and VOCs.

Each sensor has a unique sensitivity curve depending on the gas concentration. When exposed to complex organic mixtures like VOCs from cannabis, the combined response from multiple sensors creates a unique "fingerprint" for each sample [1].

2.2 Volatile Organic Compounds (VOCs) and Strain Differentiation

VOCs are low-boiling organic compounds that evaporate easily at room temperature. In cannabis, VOCs include terpenes, aldehydes, ketones, and esters, which contribute to the aroma and are strain-specific. Strains like Sativa and Indica exhibit distinct VOC profiles due to differences in terpene composition. Detecting these VOCs allows indirect strain classification using odor patterns [16],[17].

2.3 Signal Processing and ΔR Calculation

Sensor output is measured as raw analog voltage and read by an Arduino's 10-bit ADC, yielding values from 0 to 1023. To normalize readings across sessions and conditions, a differential response (ΔR) is calculated by Eq. (1):

$$\Delta R = R_{sample} - R_{baseline} \quad (1)$$

Where R_{sample} is the sample response and
$R_{baseline}$ is the average during baseline

2.4 Electronic Nose and Odor-Based Classification

An electronic nose (e-nose) mimics the human olfactory system by using multiple sensors to capture complex odor profiles. Rather than targeting specific compounds, e-noses record overall signal patterns across the array. These multidimensional data are then used to classify odors through pattern recognition techniques, making them ideal for complex mixtures like cannabis VOCs [1],[10–12].

2.5 Machine Learning Algorithms for Classification

In this study, a Random Forest algorithm was employed as the primary classification method due to its robustness, high accuracy, and ability to handle small to moderate-sized datasets effectively [18–20]. As an ensemble learning technique, Random Forest constructs multiple decision trees during training and aggregates their outputs to improve predictive performance and reduce overfitting. This method leverages diverse feature subsets and bootstrapped samples to generate uncorrelated trees, enhancing model generalization.

Compared to single Decision Trees, Random Forest offers improved stability and accuracy while preserving the interpretability of individual decision paths. This makes it a practical and scalable choice for evaluating VOC-based cannabis strain classification, particularly in sensor-driven applications.

3. Experimental Setup

3.1 Sample Description

Each cannabis sample, whether flower or leaf, was weighed prior to measurement and standardized to approximately 1 gram per sample. While the moisture content of the dried materials was not directly measured, the curing protocol aimed to reduce moisture to a typical commercial range of 10–12%, supporting consistency across replicates. All measurements were conducted in a controlled indoor environment at a room temperature of approximately 25 ± 2°C with ambient relative humidity maintained between 55–60%. The three selected cannabis strains—Black Patronas (Hybrid), MAC Gold (Indica), and Banana Daddy R1 (Sativa)—were dried and cured for three days before testing to stabilize VOC emissions. Visual representations of the dried cannabis flowers and leaves used in the experiments are shown in **Figure 1–2**, respectively.

The dried flower and leaf samples used in this study were commercially sourced from a licensed cannabis cultivator. As such, specific details regarding the exact growth stage (e.g., Day after Planting) and cultivation environment (e.g., greenhouse, outdoor, or indoor plant factory) were not disclosed to the researchers. However, all samples were visibly mature, fully cured, and consistent with commercial-grade cannabis products.

To evaluate the specificity of the sensor system, two additional non-cannabis plant samples were included: kratom (Mitragyna speciosa) leaves and tobacco (Nicotiana tabacum) leaves. These were used to analyze the sensor's ability to differentiate cannabis odors from other herbal substances.

3.2 Sensor Array and Chamber Setup

The experimental system was designed to detect and classify volatile organic compounds (VOCs) emitted from dried cannabis samples using a sensor-

based odor chamber. As shown in **Figure 3**, the setup consists of the following major components:

Figure 1 Dried flowers (**a**) Black Patronas (Hybrid) (**b**) MAC Gold (Indica) (**c**) Banana Daddy R1 (Sativa)

Figure 2 Dried leaves (**a**) Black Patronas (Hybrid) (**b**) MAC Gold (Indica) (**c**) Banana Daddy R1 (Sativa)

The bottom section housed the sample (e.g., dried cannabis flower or leaf), which was placed on a platform inside a sealed chamber. A fan was installed above the sample to circulate air and facilitate VOC flow upward. The airflow was controlled using a fan speed controller connected externally.

VOCs from the sample rose into the upper chamber, which contained an array of gas sensors (MQ-2, MQ-3, MQ-6, TGS-822, and TGS-826). Each sensor captured the VOC signature in real time. The top and bottom chambers were connected through a controllable shutter system, allowing timed exposure to the sensors.

Signal Acquisition: All sensors were connected to a printed circuit board (PCB) for signal wiring and voltage regulation. Analog signals from the PCB were transmitted to the Arduino Mega 2560 microcontroller via signal cables.

Microcontroller and Data Transmission: The Arduino collected analog voltage readings from each sensor and transmitted them to a computer via a USB connection. The system operated continuously, sending real-time sensor data to a software interface for processing.

Data Visualization and Processing: On the computer side, software such as Node-RED and Grafana was used to visualize the raw signals. The data was simultaneously stored in InfluxDB, while machine learning analysis was performed separately using Python.

The photograph of the actual experimental setup, corresponding to the schematic diagram, is shown in **Figure 4**. The chamber housed cannabis samples, with sensor arrays mounted above and controlled airflow provided by a fan and adjustable shutter. The PCB distributed sensor signals to the Arduino Mega 2560, which transmitted data to a PC via USB for further processing.

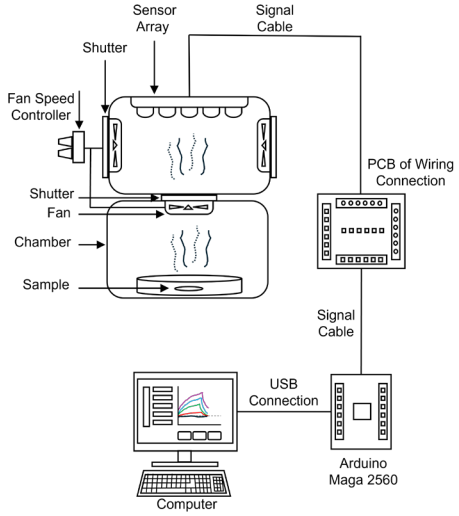

Figure 3 Schematic diagram of the sensor-based odor detection system for cannabis strain classification

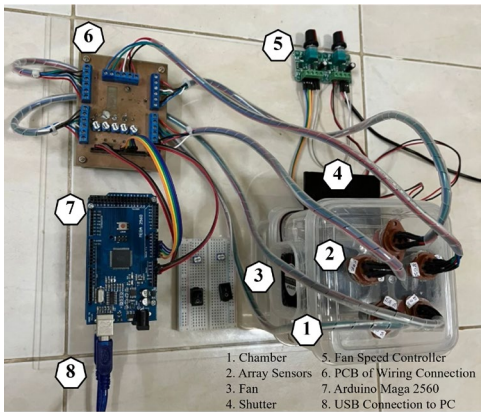

Figure 4 Sensor array and chamber setup

3.3 Data Flow: Acquisition to Visualization

The Arduino Mega 2560 collected analog voltage signals from each sensor using its built-in 10-bit ADC, resulting in values ranging from 0 to 1023. These raw analog values were transmitted via serial connection to Node-RED running on a host computer. Node-RED was configured to parse the incoming sensor data, timestamp each reading, attach strain and sample metadata, and store the data in InfluxDB, a time-series database.

Real-time data visualization was achieved using Grafana, which connected to InfluxDB and displayed live plots for each sensor channel.

Each measurement session was divided into three distinct phases to ensure consistent sensor response and reliable VOC detection:

Baseline Recording: During the initial 3 minutes, clean ambient air was circulated through the chamber to allow the sensors to stabilize. The average sensor output during this phase was recorded as the baseline value $R_{baseline}$ for each sensor.

VOC Exposure Phase (0–15 minutes): After a 3-minute baseline acquisition period, the cannabis sample (flower or leaf) was introduced into the sealed chamber by opening the shutter. A fan was activated to circulate the VOCs within the chamber, thereby

exposing the sensor array to the target odor. This exposure phase lasted for 15 minutes, during which the sensor response (ΔR) was continuously recorded.

Sensor Reset Phase (15–60 minutes): After VOC exposure, the chamber was flushed with clean air by opening both shutters and increasing the fan speed. This 45-minute flushing period allowed residual VOCs to dissipate and enabled the sensors to return to baseline before the next cycle began.

Figure 5 shows the example of sensor response pattern of the MQ-2 gas sensor to cannabis VOCs. The measurement process consisted of two main phases: exposure to the target gas (shaded region, 0–15 minutes), during which ΔR increased as VOC concentration rose, and a recovery phase with clean air (15–60 minutes), during which the sensor response gradually returned toward baseline. The differential response (ΔR) was calculated as the deviation from the baseline recorded during the initial 3-minute period.

3.4 Machine Learning Pipeline

The machine learning pipeline was designed to classify cannabis strains based on odor profiles captured by a gas sensor array. For each sample, the raw analog signals from five gas sensors (MQ-2, MQ-3, MQ-6, TGS-822, and TGS-826) were processed to compute ΔR values, representing the difference between VOC exposure and baseline conditions. These ΔR values formed the feature vector for each sample.

Each feature vector was labeled according to the cannabis strain and plant part (e.g., Hybrid flower, Indica leaf). In total, 72 labeled samples were used for model training. An additional set of 6 unseen samples-one from each of the six strain–part combinations-was reserved exclusively for testing. These test samples were not included in the training phase to ensure unbiased performance evaluation.

A Random Forest classifier was employed due to its robustness, high accuracy, and resistance to overfitting, particularly suitable for small to moderate datasets. The model was composed of 150 decision trees, each trained on a bootstrap sample of the training data, with a maximum depth of 10 to prevent over-complexity. During prediction, each tree casts a "vote," and the final classification is determined by majority rule. The decision trees in the ensemble use the Gini impurity criterion to split features that maximize class separation. While the Random Forest is an ensemble model, its underlying structure is based on interpretable Decision Trees, enabling insight into how individual features contribute to predictions.

The dataset was split using an 80/20 train-test ratio. Prior to training, features were normalized using z-score standardization. The resulting model achieved an accuracy of 0.87 on the test set, demonstrating reliable classification performance, as discussed in Section 4.6.

4. Results and Discussion

4.1 Flower vs Leaf Response

To explore the influence of plant part on gas sensor performance, comparative tests were conducted using dried flowers and leaves from three cannabis strains. **Figures 6–7** illustrate the ΔR response curves for Hybrid (Black Patronas) flowers and leaves, respectively. The MQ-2 sensor, in particular, exhibited a strong signal when exposed to flower samples, peaking at over 430 ΔR, followed by TGS-822 and TGS-826. The signal from leaf samples showed a similar temporal profile-rapid rise followed by decay—but with slightly lower peak values.

However, when examining the complete dataset summarized in **Table 1**, it became evident that this trend did not generalize across all strains. In Indica (MAC Gold) and Sativa (Banana Daddy R1), the leaf samples produced higher responses than the flowers for several sensors, notably MQ-2, TGS-822, and TGS-826. This suggested that in certain strains, the leaf tissues retained or emitted more of the sensor-detectable VOCs—such as green leaf volatiles, alcohols, or aldehydes—compared to the flower samples, especially under specific curing conditions.

These findings challenged the common assumption that cannabis flowers always yield higher VOC emissions than leaves. Instead, the VOC release profile appeared to be strain-dependent and compound-specific, influenced by both plant anatomy and the sensor sensitivity range.

The consistent non-responsiveness of MQ-6 across all samples also highlighted its limited relevance for cannabis VOC detection in this configuration.

It was important to note that the sensor response data presented in this section (Section 4.1) were obtained from samples cured on Day 2. These results reflected the influence of the curing process on VOC emission, and it was anticipated that Day 2 represented the peak response window for most cannabis strains. This assumption was further examined and discussed in Section 4.2.

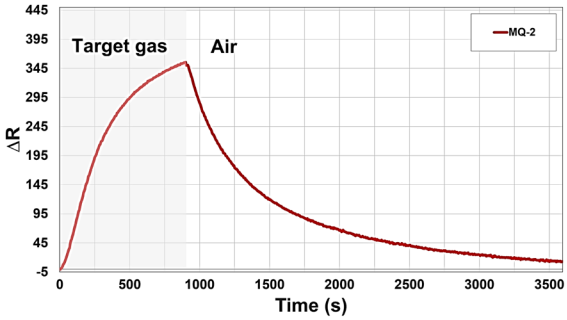

Figure 5 The example of sensor response pattern of the MQ-2 gas sensor to cannabis VOCs

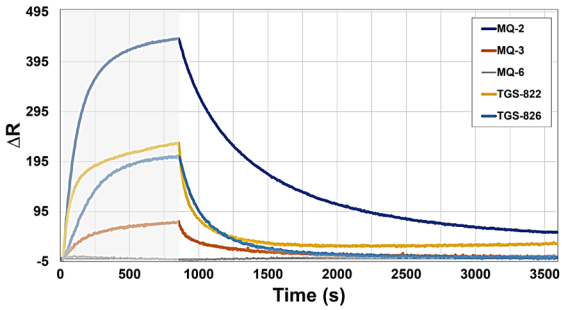

Figure 6 ΔR response curves for Hybrid (Black Patronas) flowers

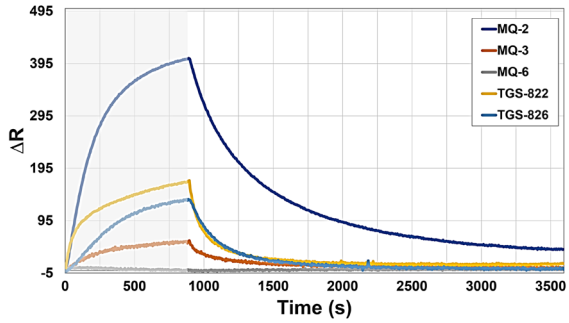

Figure 7 ΔR response curves for Hybrid (Black Patronas) leaves

Table 1 Maximum ΔR values (mean ± SD) observed from dried flower (F) and leaf (L) samples for each cannabis strain on curing Day 2.

Sensor	Hybrid (F)	Hybrid (L)	Indica (F)	Indica (L)	Sativa (F)	Sativa (L)
MQ-2	434 ± 34	363 ± 38	264 ± 53	282 ± 33	269 ± 45	286 ± 30
MQ-3	82 ± 20	50 ± 12	30 ± 5	32 ± 3	27 ± 5	38 ± 3
MQ-6	0	0	0	0	0	0
TGS-822	195 ± 31	137 ± 23	92 ± 34	103 ± 16	84 ± 31	169 ± 39
TGS-826	166 ± 27	107 ± 32	54 ± 6	54 ± 13	57 ± 23	69 ± 11

4.2 Effect of Curing Time

The curing process plays a critical role in shaping the VOC profiles emitted by cannabis plant materials. This section investigates the sensor responses over a three-day curing period for both flowers and leaves across three cannabis strains. The samples were cured under controlled indoor conditions by placing approximately 1 gram of each sample into sealed plastic containers and storing them at room temperature (approximately 25 ± 2°C) in a ventilated environment without exposure to direct sunlight. Although relative humidity was not actively regulated, it remained within an ambient range of approximately 55–60%. This protocol was applied consistently to all samples to ensure reproducibility across measurements.

As shown in **Figure 8**, among all sensors, MQ-2 consistently exhibited the strongest responses, particularly for the Hybrid flower, which peaked at Day 2 (ΔR = 434). This suggests that VOC concentrations reached their maximum around 48 hours post-curing, aligning with previous findings on terpene and aldehyde release dynamics. Interestingly, Sativa leaf also produced a strong ΔR signal on Day 1, hinting at early VOC emission in leafy tissues.

As shown in **Figures 9–10**, MQ-3 showed lower ΔR values overall but was still effective in capturing dynamic changes, with the highest value also observed at Hybrid flower—Day 2 (ΔR = 82). TGS-822, on the other hand, demonstrated strain-specific and tissue-specific responses, notably detecting Sativa leaf—Day 1 (ΔR = 181) and Hybrid flower—Day 3 (ΔR = 207), supporting its potential for temporal VOC profiling.

As shown in **Figure 11**, the TGS-826 sensor exhibited moderate responses compared to the other sensors but nevertheless captured notable differences between strains and plant parts over the three-day curing period. The Hybrid flower once again produced the strongest signal, particularly on Day 2 (ΔR = 166) and Day 3 (ΔR = 165), reinforcing the pattern of heightened VOC activity around 48-72 hours post-curing. While the overall ΔR values from Indica and Sativa samples were lower, the Sativa leaf on Day 3 (ΔR = 76) stood out as a localized VOC peak, suggesting a delayed release or compound-specific interaction detectable by TGS-826. These observations indicated that although TGS-826 was less sensitive than MQ-2, it nevertheless provided valuable supplementary information—especially for differentiating between later-stage VOC profiles.

The heatmaps confirmed that each sensor responded uniquely to the curing progression. While flowers generally emitted higher VOC levels than leaves, some exceptions (e.g., Sativa leaf) underscored the complexity of strain- and tissue-dependent VOC evolution.

These findings highlighted that Day 2 represented a crucial time window for VOC detection, particularly for strain differentiation using low-cost sensors. Future optimization of measurement timing could have enhanced classification performance and reduced mislabeling risks in real-world applications.

4.3 Sensor-specific Patterns

Each gas sensor in the array demonstrated a unique sensitivity profile to volatile organic compounds (VOCs), contributing differently to cannabis strain discrimination. To visualize these response patterns, radar charts were generated based on Day 2 ΔR values from flower and leaf samples (**Figure 12** and **Figure 13**).

The MQ-2 sensor exhibited the strongest response across all strains in both flowers and leaves, particularly in the Hybrid flower, with a ΔR of 434. This dominance was attributed to its broad sensitivity to smoke-related compounds, alcohols, and methane—indicating a high VOC concentration in Hybrid samples. Notably, the Sativa leaf also produced a high MQ-2 response (ΔR = 286), suggesting overlapping VOC profiles between some leaf and flower tissues.

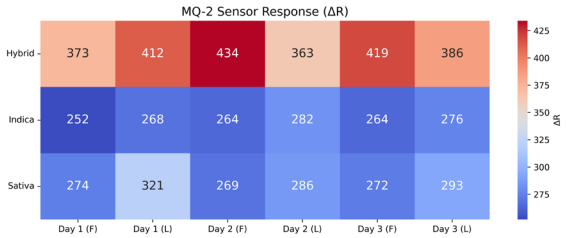

Figure 8 Heatmap of ΔR for MQ-2 across Days and Strains

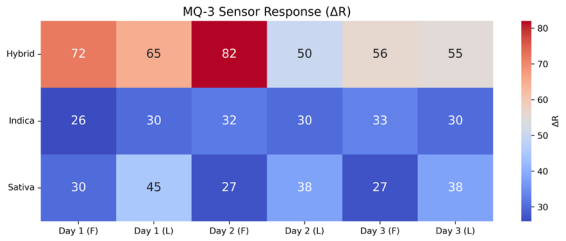

Figure 9 Heatmap of ΔR for MQ-3 across Days and Strains

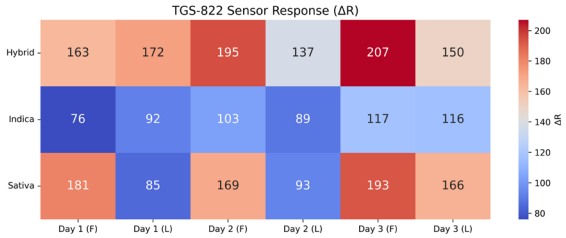

Figure 10 Heatmap of ΔR for TGS-822 across Days and Strains

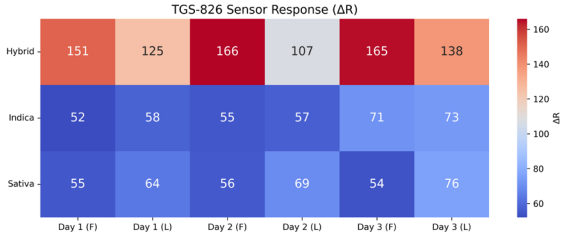

Figure 11 Heatmap of ΔR for TGS-826 across Days and Strains

MQ-3, while lower in absolute ΔR, showed distinct selectivity, especially toward alcohol-type VOCs. It responded highest to Hybrid flower (ΔR = 82) and leaf (ΔR = 50), providing a useful secondary layer for strain classification where ethanol or methanol emissions are prominent.

Among all sensors, TGS-822 stood out for its highly selective and differentiated patterns, showing peak responses in Hybrid flower (ΔR = 195) and, importantly, in Sativa leaf (ΔR = 169). This exception to the typical dominance of Hybrid responses highlighted TGS-822's strength in detecting unique VOC profiles, particularly from aromatic or alcohol-related compounds in leaves.

Figure 12 and **Figure 13** collectively demonstrate that while MQ-2 offered broad VOC detection, the complementary specificity of MQ-3 and TGS-822 enhanced the system's discriminatory power across cannabis strains and tissue types.

4.4 Notable Sensors: TGS-822

While most sensors in the array showed peak responses to the Hybrid strain, particularly in flower samples, TGS-822 stood out due to its distinct behavior with leaf samples of the Sativa strain.

As illustrated in **Figure 14**, TGS-822 displayed its highest ΔR value in response to Sativa leaf (ΔR = 217)—surpassing its response to Hybrid leaf (ΔR = 171) and Indica leaf (ΔR = 108). This pattern was not observed in any of the other four sensors, which generally favored the Hybrid strain regardless of plant part.

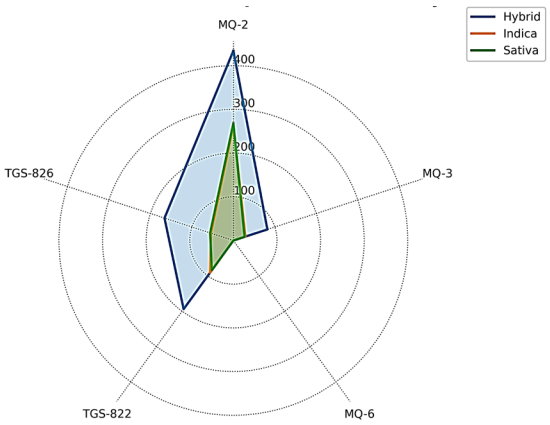

Figure 12 Radar chart of ΔR response for flower only

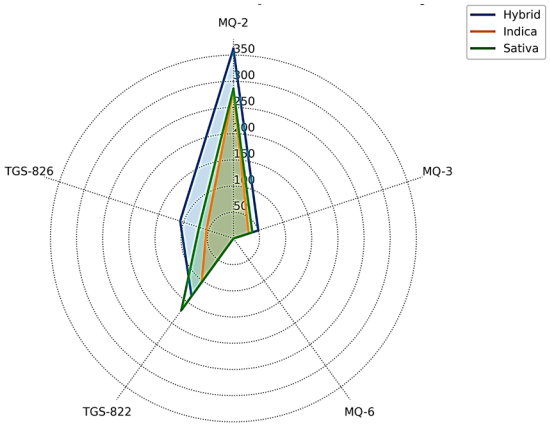

Figure 13 Radar chart of ΔR response for leaf only

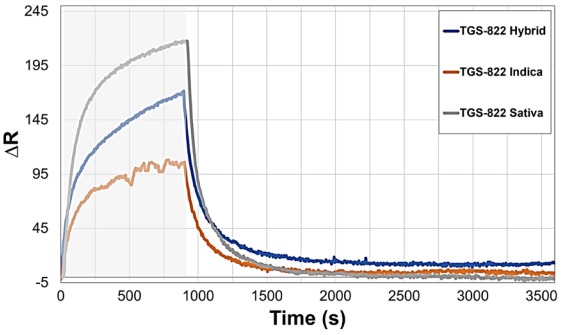

Figure 14 TGS-822 displayed its highest ΔR values in response to Sativa leaves

This exception highlighted the unique selectivity of TGS-822 toward volatile compounds likely present in the Sativa leaf, such as alcohols or aromatic compounds, for which TGS-822 was specifically designed. It suggested that the VOC profile of Sativa leaves was chemically distinct and that TGS-822 was particularly sensitive to this difference.

This divergence from the general trend underlined the importance of sensor diversity in odor-based classification. TGS-822 contributed uniquely to detecting strain-specific signatures in non-floral tissues, making it an indispensable component in a robust multi-sensor system.

4.5 VOC Comparison with Kratom and Tobacco

To evaluate the selectivity of the sensor system, non-cannabis plant materials—kratom (Mitragyna speciosa) and tobacco (Nicotiana tabacum)—were also tested using the same sensor array. The VOC profiles, expressed as ΔR values over time, were recorded and compared to those of cannabis strains.

As shown in **Figure 15**, kratom exhibited minimal sensor responses across all five sensors, with ΔR values remaining below 25 even after 900 seconds. This suggests that the VOCs emitted by kratom leaves have limited interaction with the gas sensors used, indicating low odor intensity or mismatch in chemical affinity.

In contrast, as shown in **Figure 16**, tobacco leaves triggered significantly higher responses—particularly from TGS-822 and TGS-826, reaching ΔR values over 200. These values were comparable to or even exceeded those observed in cannabis samples. This implied that tobacco shared certain VOCs or chemical structures that were detectable by the sensor array, potentially overlapping with compounds found in cannabis.

Despite these overlaps, the overall VOC fingerprint (i.e., the relative shape and dynamics of sensor responses) of kratom and tobacco differed clearly from that of cannabis strains. This suggested that a trained classification model could have effectively distinguished non-cannabis plants from cannabis, further reinforcing the selectivity and applicability of the proposed system for real-world deployment.

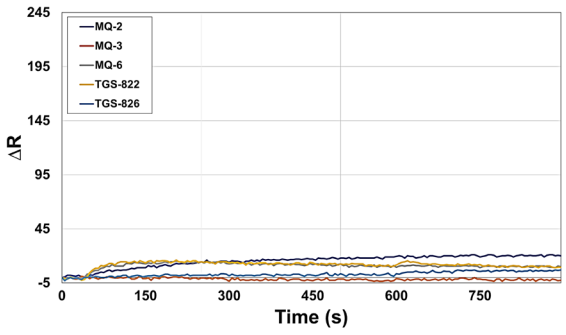

Figure 15 Kratom exhibited minimal sensor responses across all five sensors

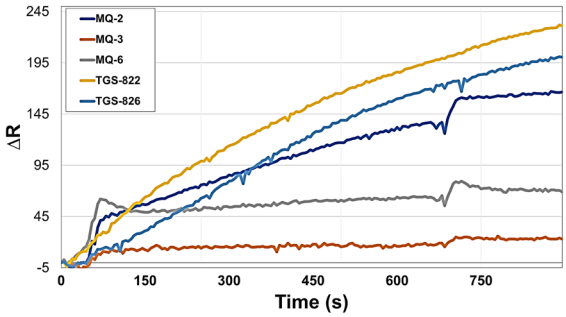

Figure 16 Tobacco leaves triggered significantly higher responses for TGS-822 and TGS-826

4.6 Strain Classification

A Random Forest classifier was used to distinguish six cannabis sample classes derived from three strains (Hybrid, Indica, Sativa) and two plant parts (flower and leaf). Each sample was represented by ΔR features from five gas sensors. All training and test samples were obtained on curing Day 2, which exhibited consistently strong sensor responses as discussed in Section 4.2.

As shown in **Table 2**, the model correctly classified 5 out of 6 test samples, achieving an accuracy of 83%. The only misclassification occurred when a Sativa flower sample was predicted as Indica flower, likely due to overlapping VOC characteristics as captured by generalist sensors such as MQ-2 or MQ-3. In contrast, leaf samples-especially Sativa leaf-were classified with higher confidence, influenced by distinctive TGS-822 and TGS-826 responses.

Although Random Forest is an ensemble method, each individual estimator is a Decision Tree, which enables interpretation of classification rules. For example, some trees first split on TGS-822 or TGS-826 to separate classes based on VOC strength or specificity. A representative decision tree is visualized in **Figure 17** to illustrate the internal structure of the ensemble.

These results highlighted the effectiveness of combining a low-cost sensor array with an interpretable machine learning approach to achieve reliable classification, even with a limited dataset. The system showed promise for further development as a scalable cannabis authentication tool based on odor profiling.

Table 2 Random Forest classification results.

	Predicted					
Actual	Hybrid (F)	Hybrid (L)	Indica (F)	Indica (L)	Sativa (F)	Sativa (L)
Hybrid (F)	1					
Hybrid (L)		1				
Indica (F)			1			
Indica (L)				1		
Sativa (F)			1			
Sativa (L)						1

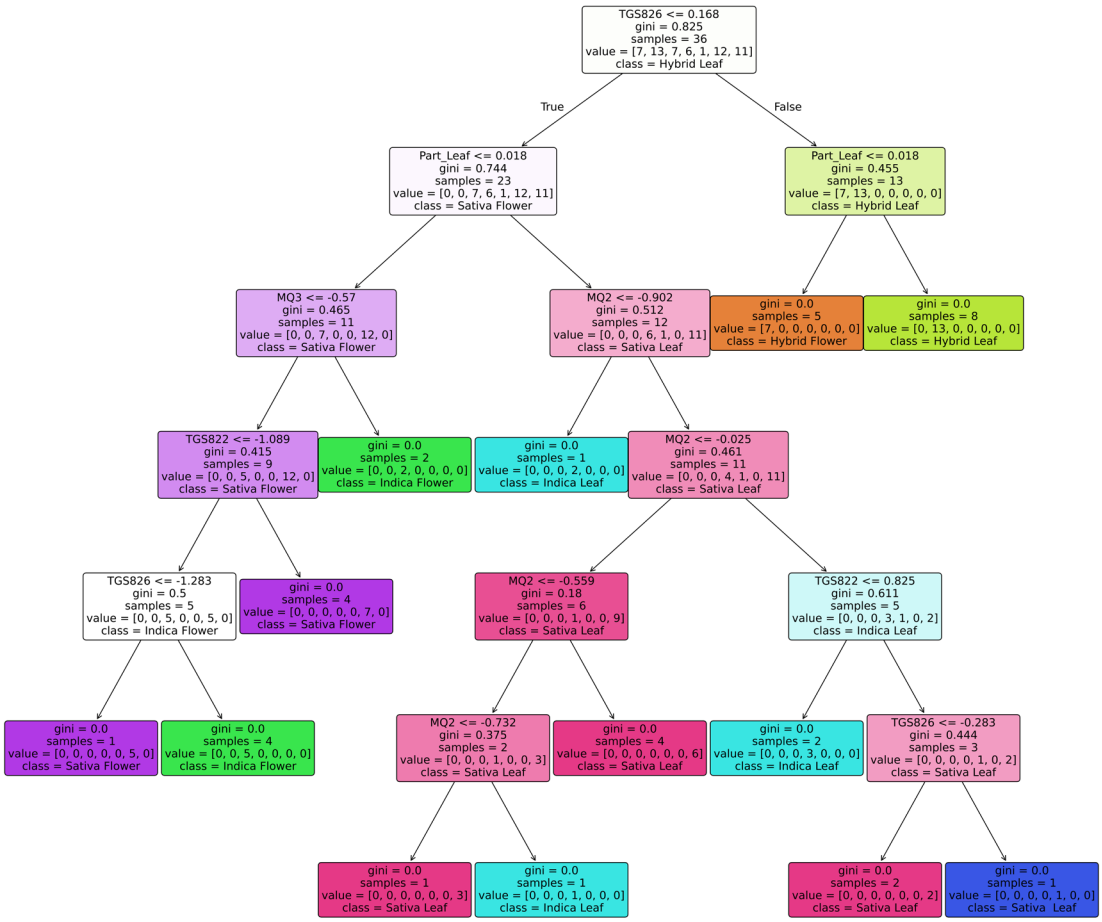

Figure 17 Example decision tree illustrating how sensor features are used to classify cannabis samples in the Random Forest model

.

The classification model based on Random Forest achieved high performance despite being trained on a relatively small dataset (72 training samples, 6 test samples). To enhance interpretability, one of the individual decision trees within the ensemble was visualized in **Figure 17**. This tree highlighted how sensor readings, particularly from TGS-822, TGS-826, and MQ-2, contributed most significantly to classification decisions. For instance, many splits occurred first on TGS-826 or TGS-822, indicating these sensors' high discriminatory power for differentiating strain–part combinations.

The relative importance of input features could also be inferred from their position in the tree hierarchy. For example, early splits on TGS-826 suggested it played a strong role in separating Hybrid from non-Hybrid samples, while MQ-2 was often used in deeper branches to refine classification between flower and leaf samples of the same strain. This hierarchical splitting aligned with known sensor sensitivities-e.g., MQ-2's broad sensitivity to VOCs and TGS-826's response to specific alcohols and esters commonly found in cannabis.

Nonetheless, limitations remained. The small dataset limited statistical power and generalizability across cultivars and growth conditions. Additionally, while Random Forest improved performance over a single decision tree, it introduced challenges in interpreting feature influence at the ensemble level. Future work should integrate feature importance metrics (e.g., Gini importance), expand the dataset across DAP stages and cultivation environments, and evaluate alternative models such as SVMs or neural networks.

5. Conclusion

This study demonstrated the feasibility of using a low-cost, multi-sensor system combined with a Decision Tree algorithm for the classification of cannabis strains based on VOC emissions. By analyzing sensor responses from both dried flowers and leaves of three commercial strains-Hybrid (Black Patronas), Indica (MAC Gold), and Sativa (Banana Daddy R1)-distinctive VOC profiles were observed. Notably, MQ-2 consistently showed the highest ΔR values for most samples, while TGS-822 uniquely responded best to Sativa leaves, highlighting its strain-specific selectivity.

The results also revealed that curing time significantly influenced VOC emissions, with Day 2

often yielding peak sensor responses. Comparisons with non-cannabis herbs (kratom and tobacco) confirmed the system's ability to differentiate cannabis odors from other botanicals. The Random Forest classifier achieved an accuracy of 83% with only a single misclassification, validating the potential of this system as a practical tool for strain verification and quality assurance in the cannabis industry.

This research provided a foundational framework for the future development of a prototype-level, portable, field-deployable cannabis authentication device using sensor arrays and machine learning.

6. Acknowledgments

This work was supported by funding from a Research Group Grant, Faculty of Engineering, Ubon Ratchathani University.

7. References

[1] Z. Haddi, A. Amari, H. Alami, N. El Bari, E. Llobet and B. Bouchikhi, "A portable electronic nose system for the identification of cannabis-based drugs," *Sensors and Actuators B: Chemical*, vol. 155, no. 2, pp. 456–463, 2011, doi: 10.1016/j.snb.2010.12.047.

[2] D. Jin, P. Henry, J. Shan and J. Chen, "Classification of cannabis strains in the Canadian market with discriminant analysis of principal components using genome-wide single nucleotide polymorphisms," *PLoS ONE*, vol. 16, no. 6, 2021, Art. no. e0253387, doi: 10.1371/journal.pone.0253387.

[3] J. M. McPartland and G. W. Guy, "Models of Cannabis Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names," *The Botanical Review*, vol. 83, no. 4, pp. 327–381, 2017, doi: 10.1007/s12229-017-9187-0.

[4] N. Herwig, S. Utgenannt, F. Nickl, P. Möbius, L. Nowak, O. Schulz and M. Fischer, "Classification of Cannabis Strains Based on their Chemical Fingerprint—A Broad Analysis of Chemovars in the German Market," *Cannabis and Cannabinoid Research*, vol. 10, no. 3, 2024, doi: 10.1089/can.2024.0127.

[5] G. Knight, S. Hansen, M. Connor, H. Poulsen, C. McGovern and J. Stacey, "The results of an experimental indoor hydroponic cannabis growing study, using the 'Screen of Green' (ScrOG) method—Yield, tetrahydrocannabinol (THC) and DNA analysis," *Forensic Science International*, vol. 202, no. 1–3, pp. 36–44, 2010, doi: 10.1016/j.forsciint.2010.04.022.

[6] J. Sawler, J. M. Stout, K. M. Gardner, D. Hudson, J. Vidmar, L. Butler, J. E. Page and S. Myles, "The Genetic Structure of Marijuana and Hemp," *PLoS ONE*, vol. 10, no. 8, 2015, Art. no. e0133292, doi: 10.1371/journal.pone.0133292.

[7] R. Zimmerleiter, W. Greibl, G. Meininger, K. Duswald, G. Hannesschläger, P. Gattinger, M. Rohm, C. Fuczik, R. Holzer and M. Brandstetter, "Sensor for Rapid In-Field Classification of Cannabis Samples Based on Near-Infrared Spectroscopy," *sensors*, vol. 24, no. 10, 2024, Art. no. 3188, doi: 10.3390/s24103188.

[8] L. Ambach, F. Penitschka, A. Broillet, S. König, W. Weinmann and W. Bernhard, "Simultaneous quantification of delta-9-THC, THC-acid A, CBN and CBD in seized drugs using HPLC-DAD," *Forensic Science International*, vol. 243, pp. 107–111, 2014, doi: 10.1016/j.forsciint.2014.06.008.

[9] L. Nahar, M. Guo and S. D. Sarker, "Gas chromatographic analysis of naturally occurring cannabinoids: A review of literature published during the past decade," *Phytochemical Analysis*, vol. 31, no. 2, pp. 135–146, 2019, doi: 10.1002/pca.2886.

[10] *MyDx Analyzer/Sensor—Electronic Nose Technology*, MyDx Inc., Mar. 2025. [Online]. Available: https://www.mydxlife.com/mydx-analyzer-sensor/

[11] H. Hendrick, H. Humaira, S. Yondri, H. Afdhal and D. Rahmatullah, "Dry Cannabis Detection by Using Portable Electronic Nose," in *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, Dec. 8–9, 2022, pp. 378–383, doi: 10.1109/ISRITI56927.2022.10053031.

[12] L. S. Leite, V. Visani, P. C. F. Marques, M. A. B. L. Seabra, N. C. L. Oliveira, P. Gubert, V. W. C. de Medeiros, J. O. de Albuquerque and J. L. L. Filho, "Design and implementation of an electronic nose system for real-time detection of marijuana," *Instrumentation Science & Technology*, vol. 49, no. 5, 2021, doi: 10.1080/10739149.2021.1887213.

[13] S. Rajora, D. K. Vishwakarma, K. Singh and M. Prasad, "CSgI: A Deep Learning based approach for Marijuana Leaves Strain Classification," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, Canada, Nov. 1–3, 2018, pp. 1–6, doi: 10.1109/IEMCON.2018.8615011 .

[14] K. Sriprateep, S. Khonjun, R. Pitakaso, T. Srichok, S. Sala-Ngam, Y. Srithep, S. Gonwirat, P. Luesak, S. Matitopanum, C. Chueadee, et al., "Hybrid Adaptive Multiple Intelligence System (HybridAMIS) for classifying cannabis leaf diseases using deep learning ensembles," *Smart Agricultural Technology*, vol. 9, 2024, Art. no. 100535, doi: 10.1016/j.atech.2024.100535.

[15] T. Islam, T. T. Sarker, K. R. Ahmed and N. Lakhssassi, "Detection and Classification of Cannabis Seeds Using RetinaNet and Faster R-CNN," *seeds*, vol. 3, no. 3, pp. 456–478, 2024, doi: 10.3390/seeds3030031.

[16] C. T. Wang, C. Wiedinmyer, K. Ashworth, P. C. Harley, J. Ortega and W. Vizuete, "Leaf enclosure measurements for determining volatile organic compound emission capacity from Cannabis spp.," *Atmospheric Environment*, vol. 199, pp. 80–87, 2019, doi: 10.1016/j.atmosenv.2018.10.049.

[17] V. Samburova, M. McDaniel, D. Campbell, M. Wolf, W. R. Stockwell and A. Khlystov, "Dominant volatile organic compounds (VOCs) measured at four cannabis growing facilities: Pilot study results," *Journal of the Air & Waste Management Association*, vol. 69, no. 11, pp. 1267–1276, 2019, doi: 10.1080/10962247.2019.1654038.

[18] I. D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," *IEEE Access*, vol. 12, pp. 86716–86740, 2024, doi: 10.1109/ACCESS.2024.3416838.

[19] A. Navada, A. N. Ansari, S. Patil and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in *2011 IEEE Control and System Graduate Research Colloquium*, Shah Alam, Malaysia, Jun. 27–28, 2011, pp. 37–42, doi: 10.1109/ICSGRC.2011.5991826.

[20] W. S. Holmes, M. P. L. Ooi, Y. C. Kuang, R. Simpkin, I. Lopez-Ubiria, A. Vidiella, D. Blanchon, G. S. Gupta and S. Demidenko, "Classifying cannabis sativa flowers, stems and leaves using statistical machine learning with near-infrared hyperspectral reflectance imaging," in *Proc. IEEE Int. Instrumentation and Measurement Technology Conf. (I2MTC)*, pp. 1–6, 2020, doi: 10.1109/I2MTC43012.2020.9129531.