

AEROVANT: A UAV-INTEGRATED MULTI - SENSOR PLATFORM FOR HAZARDOUS GAS DETECTION, THREAT LEVEL CLASSIFICATION, AND ENVIRONMENTAL MONITORING USING MACHINE LEARNING

Kein Jake Culanggo
Department of Data Science
University of Science and
Technology of Southern
Philippines - USTP CDO
Cagayan de Oro, Philippines
culanggo.keinjake012@gmail.com

Keith Laspona
Department of Data Science
University of Science and
Technology of Southern
Philippines - USTP CDO
Cagayan de Oro, Philippines
laspona.keith@gmail.com

Eurice Jean Lopez
Department of Data Science
University of Science and
Technology of Southern
Philippines - USTP CDO
Cagayan de Oro, Philippines
euricejeanlopez8@gmail.com

Chris Jallaine Mugot
Department of Data Science University of Science and
Technology of Southern Philippines - USTP CDO
Cagayan de Oro, Philippines
mugot.chrisjallaine@gmail.com

Qylle Christian Quino
Department of Data Science University of Science and
Technology of Southern Philippines - USTP CDO
Cagayan de Oro, Philippines
quinoqylle6@gmail.com

ABSTRACT

This study presents AEROVANT, a UAV-mounted electronic nose system for real-time hazardous gas detection and health risk classification in Cagayan de Oro, Philippines. The system integrates five MQ-series metal oxide sensors (butane, ethanol, LPG, hydrogen, ammonia) with environmental monitoring to overcome limitations of conventional fixed gas detection. A two-stage machine learning pipeline was implemented: sensor calibration using ensemble regression models (Random Forest, XGBoost) to convert raw signals into accurate parts-per-million concentrations, followed by a Multi-layer Perceptron neural network to classify cumulative health risks as “Stable” or “Critical” based on OSHA and ACGIH guidelines.

Calibration results show advanced nonlinear models achieved R^2 values above 0.99 for previously unstable sensors (MQ-3, MQ-9), outperforming baseline linear regression ($R^2 \sim 0.66$). The optimized MLP classifier reached 99.7% F1-score during validation, with perfect ROC-AUC and precision-recall metrics. Field deployment on 10,000 unseen samples yielded 97.54% balanced accuracy, confirming robust generalization. SHAP analysis revealed that sensors

considered “unstable” due to temperature and humidity were among the most informative, with the model capturing complex conditional relationships.

The system maintained 2.2-second inference latency on Raspberry Pi 4B and sub-4-second cloud synchronization, enabling real-time operation during 15-minute UAV flights at altitudes up to 150 meters, the Wi-Fi communication limit. These results demonstrate that low-cost sensor arrays with intelligent machine learning pipelines can achieve performance comparable to industrial-grade systems, supporting emergency response and environmental monitoring.

CCS CONCEPTS

Applied computing • Engineering Applications • Embedded Systems • Hardware • Computer Systems

KEYWORDS

Electronic Nose (E-Nose), gas leak detection, drone-based monitoring, sensor fusion and UAV

1 Introduction

Uncontrolled gas leaks from industrial facilities, storage systems, and residential equipment pose serious but often undetectable threats to human

health. Hazardous gases such as butane, ethanol, LPG, hydrogen, and ammonia can rapidly accumulate in confined or poorly ventilated environments, leading to asphyxiation, respiratory injury, or fatal exposure. Many of these gases are colorless or weakly perceptible, making early detection difficult without real-time monitoring systems.

Several commonly encountered gases present distinct yet overlapping risks. Hydrogen is highly flammable even at low concentrations, while hydrocarbons such as butane and LPG can trigger fires and explosions in enclosed spaces. Ammonia exposure causes acute irritation and respiratory damage, with concentrations above 300 ppm considered immediately dangerous to life or health. Despite regulatory limits, real-world exposure often involves mixtures of gases, where cumulative effects intensify health risks even when individual concentrations remain below threshold levels.

Globally, gas-related accidents remain a major cause of confined-space fatalities, accounting for more than 60% of such incidents. In the Philippines, repeated cases involving ammonia leaks, LPG explosions, and indoor generator use during power outages highlight the persistent danger of undetected gas exposure in both industrial and residential settings. These incidents underscore the need for proactive, mobile, and intelligent detection systems capable of operating in diverse and high-risk environments.

1.2 Objectives

The primary objective of this study is to develop AEROVANT, a portable, AI-powered electronic nose (e-nose) system capable of detecting toxic gas mixtures and assessing cumulative health risks in real time. The system is initially validated under controlled or simulated conditions to establish feasibility for future industrial monitoring and emergency response applications in Cagayan de Oro City.

Specific Objectives

1. **To Validate Methodologies for Gas Sensors:** To validate calibration methodologies for selected MQ-series gas sensors, ensuring accurate measurement of

target gases (ammonia, butane, LPG, hydrogen, and alcohol) under varying temperature and humidity conditions, thereby improving reliability for health risk assessment.

2. **To Develop a Binary Classification Model:** To develop a machine learning-based binary classification model that categorizes gas exposure as either “Critical” or “Stable,” using the ACGIH Threshold Limit Value (TLV) Mixture Formula as a reference framework. Model performance is evaluated using accuracy, precision, recall, F1-score, and false-negative rates for critical events.
3. **To Design and Develop a Multi-Sensor Toxic Gas Detection System Prototype:** To design and implement a portable multi-sensor gas detection prototype integrating MQ-series sensors and a DHT11 environmental sensor, capable of detecting toxic gases at parts-per-million (ppm) levels across typical operational temperature ranges.
4. **To Evaluate the Binary Classification Model Performance and UAV Operational Capability:** To evaluate the robustness of the classification model and the operational performance of the UAV platform in terms of sensor integration, environmental stability, payload handling, battery endurance, and overall suitability for real-world deployment.

1.3 Scope and Limitations of The Study

This study focuses on the design, development, and initial evaluation of AEROVANT, a portable AI-powered electronic nose (e-nose) system for detecting hazardous gas mixtures and classifying their cumulative health risk as either “Critical” or “Stable.” The system targets a predefined set of toxic and asphyxiating gases—butane, ethanol, liquefied petroleum gas (LPG), hydrogen, and ammonia—using a multi-sensor array composed of MQ-2, MQ-3, MQ-6, MQ-9, and MQ-135 sensors. Environmental factors affecting sensor behavior are

accounted for through temperature and humidity monitoring using a DHT11 sensor.

The scope includes sensor calibration, environmental compensation, and the implementation of a machine learning-based binary classification framework guided by the ACGIH Threshold Limit Value (TLV) Mixture Formula as a reference for cumulative exposure assessment. System validation is conducted under controlled laboratory and semi-controlled simulated environments, allowing reproducible evaluation of sensor accuracy, classification performance, and system integration with a UAV platform.

This study is limited to initial system validation and does not include full-scale deployment in uncontrolled real-world environments. Gas detection is restricted to the selected sensor suite, and potential cross-sensitivity effects inherent to MQ-series sensors are not entirely eliminated. The health risk classification is intended to support operational decision-making rather than clinical diagnosis or regulatory enforcement. Additionally, UAV evaluation is confined to basic operational feasibility, excluding long-term flight endurance, autonomous navigation in complex airspaces, and regulatory compliance considerations.

1.5 Significance of The Study

This study advances intelligent environmental hazard monitoring by demonstrating how multi-gas sensing, environmental compensation, and machine learning-based classification can be integrated into a unified, autonomous platform for real-time health risk assessment. By moving beyond single-gas, threshold-based detection, the proposed approach enables context-aware interpretation of cumulative gas exposure, producing actionable risk classifications that better reflect real-world conditions. The use of autonomous deployment further enhances safety by reducing the need for direct human exposure in hazardous environments, while real-time visualization supports rapid, evidence-based decision-making for emergency response and industrial safety. Overall, the study contributes a scalable and practical framework for improving public health protection and occupational safety, particularly in high-risk and resource-constrained settings.

2. Review of Related Literature

Hazardous gas monitoring has become increasingly important with industrial growth, urban expansion, and more frequent hazard events, as exposure to flammable (e.g., butane, LPG), asphyxiant (e.g., hydrogen), and toxic gases (e.g., ammonia, ethanol) can rapidly escalate in confined or poorly ventilated environments (Nambiar et al., 2024). Conventional fixed gas detectors, while useful in stable settings, are limited by stationary deployment and narrow spatial coverage, making them less effective in dynamic environments where gas plumes shift due to airflow, pressure changes, and irregular emission sources. As a result, UAV-mounted gas sensing systems have gained attention for their mobility and ability to collect measurements in hazardous or inaccessible areas, supporting applications such as leak localization, disaster response, and industrial safety assessments (Riviere et al., 2017; Stührenberg et al., 2024). However, the literature consistently identifies persistent barriers to reliable aerial gas sensing, including sensor cross-sensitivity, drift under changing temperature and humidity, aerodynamic interference from rotor downwash, and challenges in achieving real-time interpretation with limited onboard computing resources (Villa et al., 2016; Burgués & Marco, 2020; Szczurek et al., 2023). In terms of sensing hardware, MOS (MQ-series) sensors remain the most practical choice for lightweight multi-gas UAV platforms due to low cost and compactness, but they require calibration and compensation to address drift and environmental effects (Nambiar et al., 2024). More selective technologies such as NDIR and high-precision laser-based approaches (e.g., TDLAS) offer improved stability and selectivity but are often constrained by payload, power, and cost, limiting their suitability for portable multi-sensor UAV configurations (Kozłowski et al., 2021; Liu et al., 2022; Corbett & Smith, 2022). To improve reliability, multiple studies show that machine learning can reduce measurement error and enhance multi-gas discrimination by correcting environmental interference and modeling nonlinear sensor behavior; both deep learning (e.g., CNN/LSTM) and conventional models can achieve high classification accuracy under controlled conditions, but deep models are often too computationally heavy for edge deployment on UAVs (Bush et al., 2022; Kang et al., 2022;

Lorwongtrakool & Meesad, 2020; Attallah, 2023; Noh & Oh, 2025). Consequently, recent work emphasizes lightweight and ensemble-based methods as practical alternatives that preserve accuracy while enabling low-latency onboard inference (Mohammad, 2023; Zainottah et al., 2025). Across these studies, calibration and drift compensation are consistently treated as foundational requirements: linear regression remains a common baseline, while advanced ML-based calibration and integrated compensation strategies incorporating temperature and humidity sensors (e.g., DHT-series) improve stability and long-term performance (Rossi & Brunelli, 2017; Yulizar et al., 2023; Dong & Han, 2025; Yao et al., 2023). Overall, the literature supports UAV-based e-nose platforms as valuable for mobile hazard monitoring but highlights unresolved gaps in real-time multi-gas analysis, robust onboard calibration under variable conditions, edge-feasible analytics, and validation under complex environments where aerodynamic effects distort sampling (Burgués et al., 2021; Khan et al., 2024; Stührenberg et al., 2024), motivating integrated systems that combine multi-sensor fusion, environmental compensation, and efficient onboard machine learning for actionable risk assessment.

3. Methodology

This study adopted an end-to-end data-driven engineering methodology that integrates gas sensing, machine learning, embedded systems, and UAV deployment into a unified operational pipeline. The methodology was designed to transform raw gas sensor signals into actionable health risk classifications through systematic calibration, modeling, and real-time deployment.

The overall framework follows a two-stage modeling strategy:

- (1) *Sensor calibration, where raw analog sensor readings are converted into gas concentration estimates (ppm), and*
- (2) *Hazard classification, where calibrated concentrations are interpreted to determine environmental risk levels.*

This pipeline was implemented iteratively across seven phases—data collection, calibration, operational data acquisition, preprocessing, model training and optimization, system integration, and testing—ensuring robustness, reproducibility, and deployment feasibility. An overview of the complete pipeline is illustrated in Figure 3-1, with phase objectives summarized in Table 3-1.

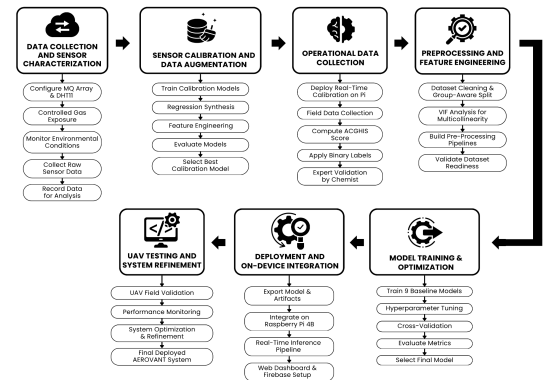


Figure 3-1. The Data Science Pipeline for the Drone Mounted E-Nose

3.1 Phase I – Sensor Data Acquisition and Characterization

3.2.1 Sensor Configuration. The electronic nose system consisted of five metal-oxide semiconductor (MOS) gas sensors—MQ-2 (Butane), MQ-3 (Ethanol), MQ-6 (LPG), MQ-9 (Hydrogen), and MQ-135 (Ammonia)—each assigned to a single target gas to minimize cross-sensitivity effects. These sensors operate by measuring resistance changes in a heated tin dioxide (SnO_2) layer when exposed to gas molecules.

Because MOS sensors are highly sensitive to environmental conditions, ambient temperature and relative humidity were continuously measured using a DHT11 sensor. All sensor outputs were digitized using an external analog-to-digital converter and logged alongside environmental metadata to support subsequent calibration and compensation.

3.2.2 Controlled Gas Exposure and Logging. Initial data collection was conducted under laboratory

conditions using predefined gas concentrations ranging from 100 ppm to 1000 ppm. Sensors underwent a stabilization period to establish baseline resistance before exposure. During each exposure cycle, raw ADC values, temperature, humidity, and timestamps were recorded at fixed intervals.

This phase produced structured datasets that captured nonlinear sensor response behavior, environmental dependencies, and sensor noise characteristics—providing the empirical foundation for calibration modeling.

3.3 Phase II – Sensor Calibration and Environmental Compensation

3.3.1 Calibration as a Supervised Regression Problem. Sensor calibration was formulated as a supervised regression task, where raw ADC readings and environmental variables were mapped to known gas concentrations in parts per million (ppm). The predictive function can be expressed as:

$$\hat{C} = f(\text{ADC}, T, H) \quad (1)$$

3.3.2 Calibration Models. Three regression models were evaluated:

- **Linear Regression (Baseline):** Provided interpretability and served as a reference for assessing nonlinearity.
- **Random Forest Regressor:** Captured nonlinear interactions and reduced overfitting through ensemble averaging.
- **XGBoost Regressor:** Offers high accuracy and robustness via gradient boosting and regularization.

Calibration performance was assessed using MAE, RMSE, and R^2 metrics, and the most stable model for each sensor was selected for deployment.

3.3.3 Data Augmentation and Feature Engineering. To improve generalization beyond discrete calibration levels, regression-based data augmentation was applied to synthesize intermediate gas concentrations. Interaction features combining ADC values with temperature

and humidity were introduced to explicitly model environmental effects. These steps produced dense and environmentally representative calibration datasets.

3.4 Phase III – Operational Data Collection

3.4.1 Multi-Environment Sampling. Following calibration, real-time data collection was conducted in laboratory, simulated, and outdoor environments. Calibrated models were embedded into the onboard system, enabling immediate conversion of sensor readings into ppm values during operation. Each data record included calibrated gas concentrations, environmental parameters, timestamps, and collection identifiers to preserve temporal and contextual integrity.

3.5 Phase IV – Composite Risk Scoring and Label Generation

3.5.1 Additive Concentration of Gas Hazard Index Score (ACGHIS). To represent cumulative exposure risk from multiple gases, a composite hazard indicator—the Additive Concentration of Gas Hazard Index Score (ACGHIS)—was computed for each observation using established occupational exposure guidelines:

$$ACGHIS = \sum_{i=1}^n \frac{C_i}{L_i} \quad (2)$$

3.5.2 Label Assignment. Observations were labeled as:

- **Stable** if $ACGHIS < 1$
- **Critical** if $ACGHIS \geq 1$

These labels served as ground truth for supervised learning. Importantly, ACGHIS was used only for labeling and validation, not as an input feature for classification models.

All exposure limits, computations, and labels were independently reviewed and validated by licensed chemical experts to ensure regulatory accuracy and ethical compliance.

3.6 Phase V – Hazard Classification Modeling

3.6.1 Problem Formulation. Hazard classification was defined as a binary supervised learning problem, predicting whether an observed environment is Stable or Critical based on calibrated gas concentrations and environmental features.

3.6.2 Model Training and Optimization. Multiple classifiers—including Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machines, Neural Networks, and XGBoost—were trained and evaluated. Group-aware cross-validation was employed using collection identifiers to prevent data leakage.

Hyperparameter optimization was performed using Bayesian search (Optuna), with F1-score as the primary objective metric to balance false alarms and missed hazards.

3.7 Phase VI – Two-Way Pipeline Validation

The methodology explicitly incorporated a two-way validation mechanism:

- **Forward Direction:** Sensor readings → calibration → ppm values → ACGHIS → expert-validated labels.
- **Reverse Direction:** ML predictions → comparison against ACGHIS-derived labels → performance validation.

This ensured that machine learning outputs remained consistent with chemical safety principles while retaining adaptability to empirical data patterns.

3.8 Phase VII – System Integration and UAV Deployment

The finalized calibration and classification pipeline was deployed on a Raspberry Pi 4B mounted on a UAV platform. Onboard inference enabled real-time hazard classification without reliance on continuous cloud computation. Data was transmitted to a cloud backend for visualization via a web-based dashboard.

System performance was evaluated in terms of sensing accuracy, classification reliability, inference latency, power consumption, and flight stability. Iterative refinements were applied to sensor placement, sampling intervals, and software efficiency based on field observations.

4 Results and Discussion

4.1 Sensor Response

4.1.1 Sensor Response Behavior. *MQ2 Sensor Response Characterization:*

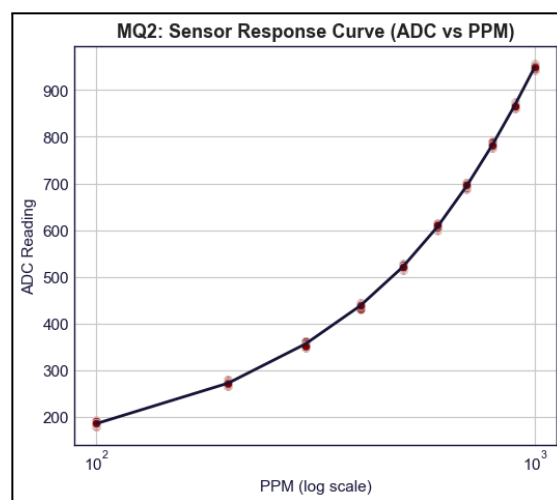


Figure 4-1. MQ2 - Butane Sensor Response Curve

The MQ sensor exhibits a clear, monotonic nonlinear relationship between its raw ADC output and gas concentration, consistent with expected chemiresistive sensing behavior. The response curve is smooth and well-defined across the entire tested concentration range, indicating predictable sensitivity increases as gas levels rise. Measurements show very low dispersion around the fitted curve, demonstrating high repeatability and a strong signal-to-noise ratio. No concentration

range exhibits abnormal variability or instability, suggesting that the sensor response is governed by a stable physical process rather than random noise or environmental interference.

This stability significantly simplifies calibration, as model development can focus on accurately capturing the sensor's inherent nonlinearity rather than compensating for drift or erratic behavior. As a result, nonlinear regression models are sufficient to reliably map ADC readings to gas concentration values, confirming the MQ sensor's suitability for precise and consistent gas measurement.

4.1.2 Sensor Response Behavior: MQ3 Sensor Response Characterization:

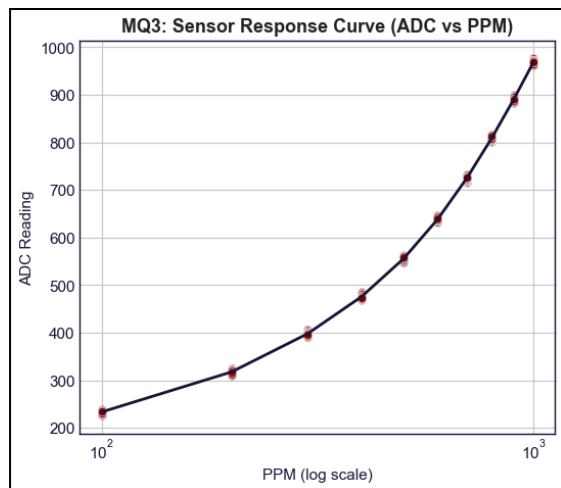


Figure 4-2. MQ3 - Ethanol Sensor Response Curve

The MQ-3 ethanol sensor demonstrates a clear monotonic and nonlinear relationship between its raw ADC output and ethanol concentration, forming a smooth convex response curve across the tested range. This consistent trend confirms the sensor's ability to track changes in ethanol concentration and provides a predictable basis for calibration.

Unlike the highly stable response observed in the MQ-2 sensor, the MQ-3 exhibits a noticeable increase in measurement variability as concentration rises. At lower ethanol levels, readings remain tightly clustered around the fitted curve, whereas at higher concentrations the data spread widens significantly, indicating that sensor

variance increases with signal magnitude. This behavior reflects concentration-dependent uncertainty, where measurement precision decreases as the sensor approaches higher exposure levels.

From a calibration standpoint, this non-uniform variance limits predictive reliability at elevated concentrations. While nonlinear regression can still capture the average response trend, calibration accuracy degrades in the upper range, requiring that high-concentration ethanol measurements be interpreted with greater uncertainty. This characteristic highlights the need for variance-aware modeling or cautious interpretation when using the MQ-3 sensor for high-level ethanol detection.

4.1.3 Sensor Response Behavior: MQ6 Sensor Response Characterization:

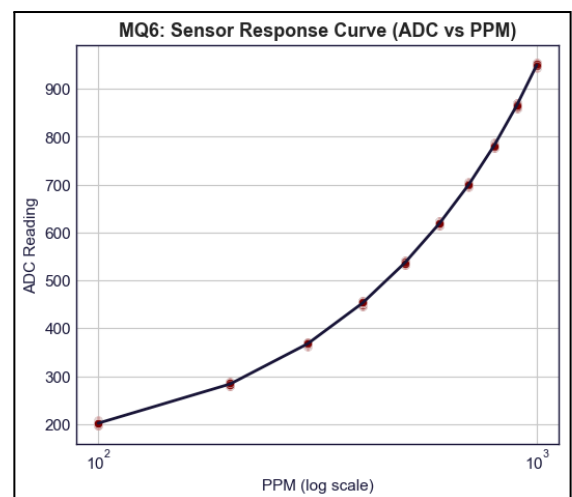


Figure 4-3. MQ6 - LPG Sensor Response Curve

The MQ-6 sensor demonstrates a stable and well-defined nonlinear response to LPG concentration, with ADC values increasing smoothly across the tested range. The response remains tightly aligned with the fitted curve from low to high concentrations, indicating strong sensitivity and predictable behavior.

A key distinguishing feature of the MQ-6 is the uniformity of its measurement precision. Unlike sensors that exhibit increasing variability at higher concentrations, the MQ-6 maintains consistently

low scatter throughout its operating range. This indicates a stable sensing mechanism with minimal noise amplification or saturation effects, even as concentrations approach upper limits.

This consistent response simplifies calibration and supports the use of standard nonlinear regression without additional variance correction. As a result, the MQ-6 enables reliable and high-precision LPG quantification across its full dynamic range, reinforcing its suitability for accurate gas concentration measurement.

4.1.4 Sensor Response Behavior: MQ9 Sensor Response Characterization:

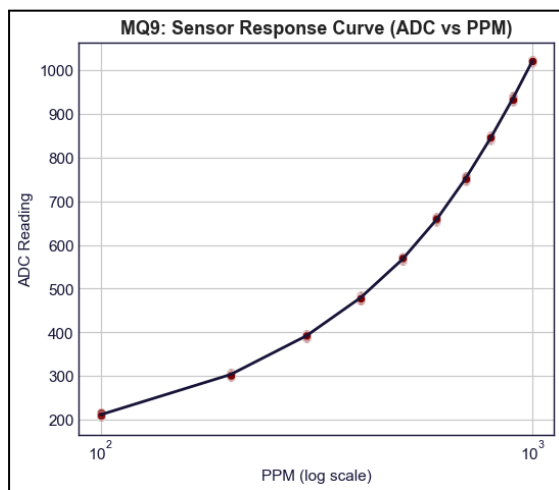


Figure 4-4. MQ9 - Hydrogen Sensor Response Curve

The MQ-9 sensor exhibits a clear nonlinear and monotonic response to hydrogen concentration, with ADC values increasing smoothly across the tested range. This confirms its fundamental sensitivity and establishes a reliable base trend for calibration.

Unlike the more stable MQ-2 and MQ-6 sensors, the MQ-9 shows a pronounced increase in measurement variability as concentration rises. While readings remain tightly clustered at lower hydrogen levels, the dispersion of ADC values expands noticeably at higher concentrations, indicating concentration-dependent uncertainty. This behavior reflects reduced signal stability under high gas exposure and greater susceptibility to amplified noise or environmental influence.

As a result, calibration of the MQ-9 cannot assume uniform measurement precision across its range. Models that do not account for this variance inflation risk overstating accuracy at higher concentrations. Effective calibration therefore requires variance-aware approaches, such as weighted regression, to ensure reliable hydrogen quantification under elevated exposure conditions.

4.1.5 Sensor Response Behavior: MQ135 Sensor Response Characterization:

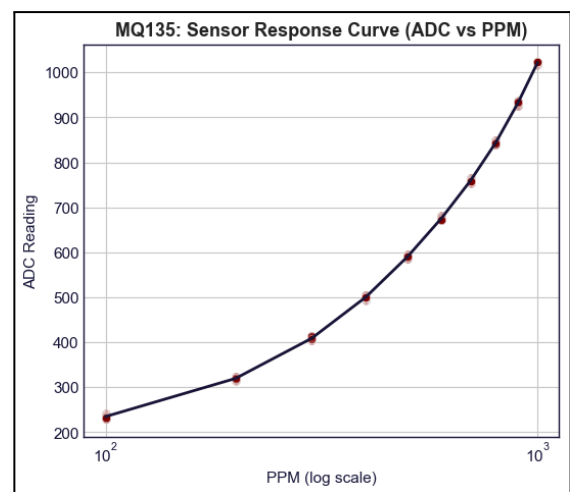


Figure 4-5. MQ9 - Ammonia Sensor Response Curve

The MQ-135 ammonia sensor demonstrates a stable and well-defined nonlinear response, with ADC readings increasing predictably across the tested concentration range. Measurements remain tightly clustered around the fitted curve, showing minimal and uniform variance throughout, indicating strong signal fidelity and consistent precision.

This stable behavior simplifies calibration, as standard nonlinear regression is sufficient to accurately model the sensor response without additional variance correction. Overall, the response analysis highlights that while all MQ sensors exhibit nonlinear trends, their reliability differs markedly: MQ-2, MQ-6, and MQ-135 show stable, high-fidelity behavior, whereas MQ-3 and MQ-9 require variance-aware modeling due to concentration-dependent uncertainty.

4.2 Quantitative Correlation Analysis

Following the baseline response curve characterization, a quantitative correlation analysis was conducted to examine the linear relationships between raw sensor outputs and ambient environmental variables—temperature and relative humidity. This analysis aimed to determine whether simple, first-order environmental drift could explain observed variations in sensor behavior. Across all MQ-series sensors (MQ-2, MQ-3, MQ-6, MQ-9, and MQ-135), the Pearson correlation matrices (Figures 4-6 to 4-10) reveal a consistent and notable result: the raw ADC outputs exhibit near-zero linear correlation with both temperature and humidity. For all sensors, correlation coefficients between ADC and temperature ranged approximately from -0.05 to -0.08 , while correlations with relative humidity ranged from about 0.03 to 0.07 . These values indicate statistical independence under linear assumptions.

A strong inverse correlation between temperature and humidity (≈ -0.85) was consistently observed across all datasets. This relationship reflects normal physical behavior of air under controlled or semi-controlled conditions and confirms the internal consistency of the environmental measurements. Importantly, however, this coupling between environmental variables did not translate into a direct linear influence on sensor outputs. This finding carries critical implications. For sensors previously identified as stable—MQ-2, MQ-6, and MQ-135—the lack of linear

environmental correlation explains their high signal fidelity and uniform variance. Their raw outputs are not measurably perturbed by ambient temperature or humidity changes in a linear sense, allowing calibration to focus primarily on modeling intrinsic sensor nonlinearity.

More importantly, the same lack of correlation was observed for MQ-3 and MQ-9, despite these sensors exhibiting pronounced heteroscedasticity at higher gas concentrations. This decisively rules out simple linear environmental drift as the cause of their instability. If temperature or humidity were directly responsible, measurable linear correlations would be expected. Their absence implies that the observed noise amplification arises from more complex, concentration-dependent mechanisms, such as sensor saturation effects or nonlinear interactions between gas concentration and environmental conditions.

Overall, the correlation analysis demonstrates that environmental effects on MQ sensors cannot be adequately described by linear relationships alone. This insight directly motivates the need for nonlinear modeling strategies and interaction-aware calibration approaches, particularly for sensors exhibiting concentration-dependent variance.

4.2.1 Detailed Analysis of Environmental Influence on Sensor Behavior. While the correlation matrices establish the absence of global linear dependence, a more granular examination of sensor response under varying environmental conditions reveals a different and complementary insight. Figures 4-11 through 4-15 illustrate how temperature and humidity systematically influence sensor outputs when examined conditioned on gas concentration levels.

Across all MQ sensors, a consistent directional pattern emerges:

- **Temperature exhibits a negative relationship with ADC output**, where increasing temperature leads to a reduction in sensor signal.
- **Relative humidity exhibits a positive relationship with ADC output**, where

increasing humidity amplifies the sensor signal.

This opposing influence is observed consistently for MQ-2, MQ-3, MQ-6, MQ-9, and MQ-135, indicating that temperature and humidity act as secondary modifiers of sensor response rather than primary drivers.

From a physical standpoint, the temperature-induced signal suppression is consistent with thermally driven desorption effects, where elevated temperatures reduce the residence time of gas molecules on the sensing surface, lowering steady-state conductivity. Conversely, the humidity-induced signal amplification is characteristic of metal-oxide semiconductor behavior, where adsorbed water molecules contribute charge carriers or alter surface conductivity, artificially inflating sensor output and partially mimicking reducing gases. The apparent contradiction between this behavior and the earlier correlation analysis is resolved by recognizing that these environmental effects are conditional rather than global. Pearson correlation measures average linear dependence across the entire dataset; it is insensitive to effects that emerge only under specific operating regimes—such as high gas concentrations. Thus, temperature and humidity do influence sensor readings, but their impact is nonlinear and concentration-dependent, explaining why simple correlation coefficients remain near zero.

This distinction is particularly important for MQ-3 and MQ-9. For these sensors, environmental modulation becomes increasingly pronounced at higher analyte loads, compounding intrinsic sensor instability and contributing to heteroscedastic error profiles. In contrast, MQ-2, MQ-6, and MQ-135 maintain stable variance despite environmental modulation, indicating stronger intrinsic robustness.

4.3 Calibration Model Performance

Linear Regression was trained to predict gas concentration (ppm) from raw ADC, temperature, and humidity. Table 4-1 summarizes the comparative performance metrics averaged across all five sensors. This is the baseline model.

Table 4-1. Comparative Performance of Linear Regression Across All Sensors

Sensor	MAE	RMSE	R- Squared Score
MQ-2	22.64	33.74	0.986354
MQ-3	101.91	167.13	0.665256
MQ-6	25.36	37.67	0.982990
MQ-9	103.70	168.32	0.660490
MQ-135	28.074149	41.37	0.979484

Rather than serving as a final modeling solution, the Linear Regression results function as a baseline stress test of sensor behavior under minimal modeling assumptions.

The strong performance observed for MQ-2, MQ-6, and MQ-135 indicates that these sensors operate within a regime where their nonlinear physical response can be locally approximated by a linear mapping without substantial loss of accuracy. The high R^2 values and low error metrics suggest that, for these sensors, environmental influences and intrinsic sensor noise do not significantly distort the concentration–signal relationship. This reinforces earlier findings that these sensors exhibit homoscedastic error structures and stable response curves, allowing simpler models to remain effective.

In contrast, the failure of Linear Regression for MQ-3 and MQ-9 is not merely a matter of reduced accuracy but a structural modeling mismatch. The low R^2 values and large errors indicate that substantial, systematic information in the data remains unexplained under linear assumptions.

When interpreted alongside the response curve and correlation analyses, this breakdown reflects the presence of concentration-dependent variance and complex interaction effects rather than random noise. Linear models, by design, assume constant error variance and additive relationships—assumptions that are explicitly violated by the heteroscedastic behavior observed in these sensors.

This divergence has important methodological implications. It demonstrates that calibration performance is fundamentally sensor-dependent and that a uniform modeling strategy is inappropriate for heterogeneous sensing arrays. While linear calibration may suffice for stable sensors, applying the same approach to MQ-3 and MQ-9 would result in misleading confidence and unreliable predictions, particularly at higher concentration levels where accurate estimation is most critical for safety applications.

4.3.1 Comparative Results of Calibration Models Across All Sensors. This section synthesizes the calibration performance across all developed models, moving beyond baseline linear analysis to a direct comparison of Linear Regression, Random Forest, and XGBoost across all five sensors. The goal is to identify the most accurate and robust calibration strategy and to test the hypothesis that while simpler models suffice for stable sensors (MQ-2, MQ-6, MQ-135), unstable sensors (MQ-3, MQ-9) require nonlinear, interaction-aware methods to address heteroscedasticity and complex dependencies. The results quantify the performance gains achieved by advanced machine learning models over linear calibration.

Table 4-2. MQ2_Butane Calibration Results

Model	MAE	RMSE	R- Squared Score
Linear Regression	22.64	33.74	0.9863
Random Forest Regressor	16.03	27.34	0.9910
XGBoost Regressor	16.98	27.52	0.9909

The results strongly reinforce the MQ-2's classification as a "stable" sensor. The baseline Linear Regression model performs exceptionally well on its own, achieving a high R2 of 0.986 and an MAE of 22.64. This indicates that the sensor's response is already highly linear and predictable.

Table 4-3. MQ3_Ethanol Calibration Results

Model	MAE	RMSE	R- Squared Score
Linear Regression	101.91	167.13	0.6652
Random Forest Regressor	20.11	34.18	0.9859
XGBoost Regressor	20.49	34.02	0.9861

The baseline Linear Regression model utterly fails to calibrate the MQ-3 sensor, as evidenced by its extremely poor performance: a low R-Squared (R2) of 0.665 and a massive Mean Absolute Error (MAE) of 101.92 PPM. This result confirms that the sensor's response is deeply nonlinear and that its significant heteroscedasticity and environmental interactions cannot be resolved by a simple linear model.

Table 4-4. MQ6_LPG Calibration Results

Model	MAE	RMSE	R- Squared Score
Linear Regression	25.36	37.67	0.9829
Random Forest Regressor	17.52	30.08	0.9891
XGBoost Regressor	18.74	30.53	0.9888

The validation metrics for the MQ-6 sensor, as summarized in Table 4-4, confirm its

classification as a "stable" sensor, exhibiting behavior very similar to the MQ-2. Its fundamentally linear and predictable nature is evident in the strong baseline performance of the Linear Regression model, which achieved a high R-Squared (R^2) of 0.983 and a respectable Mean Absolute Error (MAE) of 25.36.

Table 4-5. MQ9_Hydrogen Calibration Results

Model	MAE	RMSE	R- Squared Score
Linear Regression	103.70	168.32	0.6604
Random Forest Regressor	16.87	29.08	0.9898
XGBoost Regressor	17.60	29.47	0.9895

In sharp contrast, the machine learning models deliver a dramatic improvement. The Random Forest model, which emerged as the top performer, achieves an R-squared of 0.99 and a mean absolute error of only 16.87. This represents an 84 percent reduction in mean error compared to the linear model. The XGBoost model performs almost identically, further reinforcing the reliability of this result. The success of these models lies in their ability to utilize the engineered interaction features—such as combinations between sensor output, temperature, and humidity—to learn and correct for the nonlinearities that caused the linear model to fail.

Table 4-6 MQ135_Ammonia Calibration Results

Model	MAE	RMSE	R- Squared Score
Linear Regression	28.07	41.37	0.9794
Random Forest Regressor	19.06	32.64	0.9872
XGBoost Regressor	20.32	33.05	0.9869

This outcome mirrors the pattern observed in other stable sensors such as the MQ-2 and MQ-6. It indicates that while the MQ-135's response is fundamentally stable, small nonlinear effects influenced by environmental conditions still exist. The tree-based models effectively captured and corrected for these subtle interactions, providing a finer level of calibration precision. Overall, these results validate the role of machine learning not only as a corrective solution for unstable sensors but also as a powerful enhancement tool for stable ones—refining accuracy and ensuring optimal reliability in environmental measurement.

4.4 Bias and Scale Correction

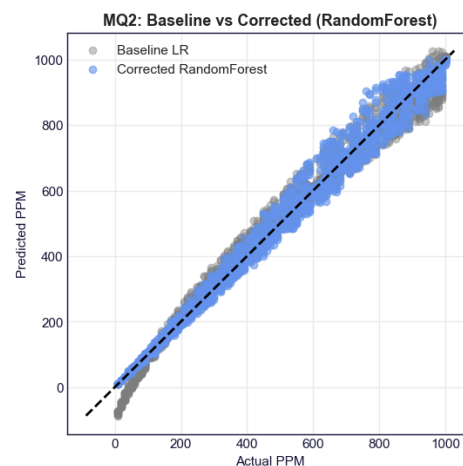


Figure 4-6. MQ2 - Baseline and Corrected Best Model Prediction Comparison

The corrected Random Forest model not only minimizes random error but also removes systematic bias, resulting in predictions that are both stable and trustworthy across all operating conditions. This final plot serves as the conclusive visual validation that the two-stage calibration framework—advanced modeling followed by targeted post-correction—achieves optimal sensor reliability and performance.

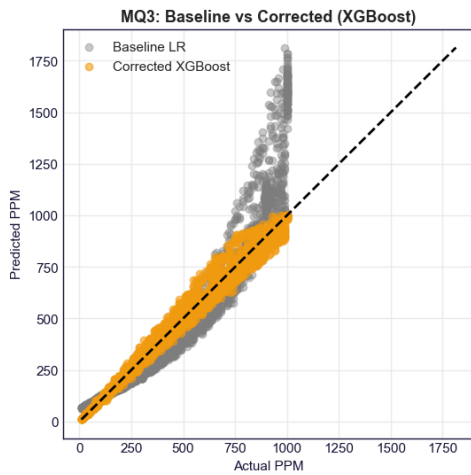


Figure 4-7. MQ3 - Baseline and Corrected Best Model Prediction Comparison

The baseline model's failures are immediately evident. At higher concentrations, predictions explode with extreme variance, reaching up to 1750 PPM. This behavior reflects the sensor's inherent nonlinear response and saturation effects, which a linear model cannot capture, and is consistent with the catastrophic evaluation metrics (MAE ~ 101.9 , $R^2 \sim 0.665$).

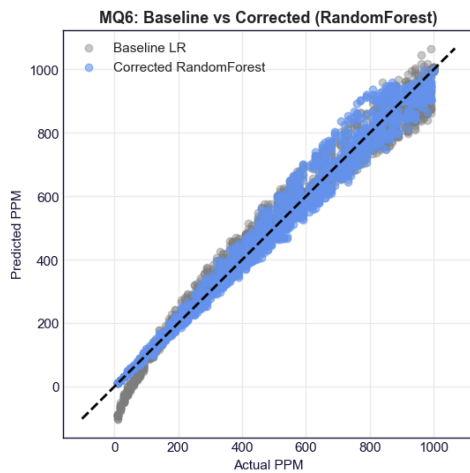


Figure 4-8. MQ6 - Baseline and Corrected Best Model Prediction Comparison

The final corrected model is superior in both precision and accuracy, validating the pipeline's effectiveness and its general applicability for sensors across different stability profiles. The

MQ-6 outputs are now robust, unbiased, and highly reliable for practical deployment.

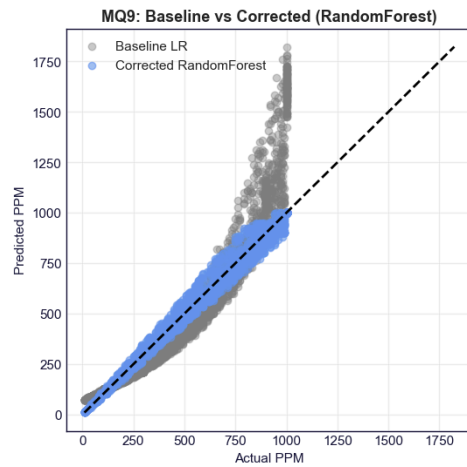


Figure 4-9. MQ9 - Baseline and Corrected Best Model Prediction Comparison

In some cases, predictions reach 1750 PPM when the actual gas concentration is only 1000 PPM. This visual pattern mirrors the catastrophic evaluation metrics (MAE ≈ 103.7 , $R^2 \approx 0.660$) and confirms that a simple linear approach is entirely inadequate for this sensor.

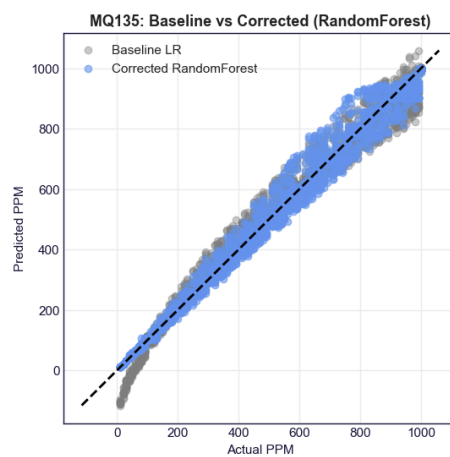


Figure 4-10. MQ135 - Baseline and Corrected Best Model Prediction Comparison

This figure serves as the final visual confirmation of the MQ-135 calibration. Even for a stable sensor with a strong baseline fit, the two-stage strategy delivers meaningful improvements. The final corrected model is both highly precise and

unbiased, validating the effectiveness and broad applicability of this methodology across all sensor types.

4.5 Classification Phase

The dataset consists of ~140,000 observations grouped into ~7,000 single-label collection sessions, dominated by real-world outdoor and laboratory data, ensuring diversity and strong data integrity with no missing values. Class distribution is moderately imbalanced (~59% Stable, 41% Critical), necessitating recall- and F1-focused evaluation. Sensor distributions show that high readings reliably indicate Critical events, while low to mid-range values across all sensors form overlapping “ambiguous zones” where single-feature decisions fail. Stable sensors (MQ-2, MQ-6, MQ-135) exhibit clean baselines with clear critical tails, whereas unstable sensors (MQ-3, MQ-9) show broad, heteroscedastic critical distributions. Temperature and humidity provide useful but non-decisive context. Overall, the results confirm the need for multivariate, context-aware modeling that integrates sensor and environmental features to resolve ambiguity and reduce false negatives.

Table 4-7. Detailed Class Distribution

Class	Count	Percentage
Stable	82,620	59.01%
Critical	57,380	40.99%

4.5.1 Baseline Performance. The cross-validated

baseline results (Table 4-8) reveal exceptionally high performance across all models, with F1 scores exceeding 0.945. This confirms that the engineered feature set—including sensor readings, temperature, humidity, and interaction terms—captures a clear and separable signal for distinguishing between “Stable” and “Critical” conditions. The data is therefore well-prepared for machine learning, with minimal noise obscuring the target patterns.

Table 4-8. Detailed Results Across All Models - Baseline

Model	Validated Train (Accuracy)	Validated (F1)
MLP	0.964143	0.965029
Logistic Regression	0.940795	0.945119
SVM	0.964652	0.965451
Gradient Boosting	0.963161	0.964197
AdaBoost	0.953866	0.956368
XGBoost	0.963295	0.964301
Random Forest	0.962580	0.963698
KNN	0.959688	0.961250
Decision Tree	0.953562	0.956041

4.5.2 Optimized Model Final Parameters. To enhance baseline performance, each model underwent a systematic hyperparameter optimization phase using Bayesian optimization with 50 independent trials. This process iteratively explored the parameter space to identify configurations that maximized validation

performance while maintaining robust generalization. Table 4-9 summarizes the optimal hyperparameters discovered for all nine models.

Table 4-9. Optimal Hyperparameters for Each Machine Learning Model

Model	Best Parameters Identified After Tuning
MLP	hidden_layer_sizes = (50, 50), activation = <i>tanh</i> , solver = <i>adam</i> , alpha = 0.0005355, learning_rate_init = 0.0001487
Logistic Regression	Solver = <i>lbfgs</i> , Penalty = <i>l2</i> , C = 2.1248, Max Iter = 839, Tolerance = 0.0007959
SVM	C = 0.00158, kernel = <i>poly</i> , degree = 5, gamma = <i>auto</i>
Gradient Boosting	n_estimators = 475, learning_rate = 0.0406, max_depth = 3, subsample = 0.5007
AdaBoost	n_estimators = 473, learning_rate = 0.9129
XGBoost	n_estimators = 334, learning_rate = 0.2842, max_depth = 2, subsample = 0.9735, colsample_bytree = 0.8215, gamma = 2.9510, reg_lambda = 9.6680
Random Forest	n_estimators = 162, max_depth = 24, max_features = <i>log2</i> , min_samples_split = 8, min_samples_leaf = 1, bootstrap = <i>False</i>
KNN	n_neighbors = 30, weights = <i>distance</i> , p = 1 (Manhattan distance), algorithm = <i>auto</i>
Decision Tree	criterion = <i>log_loss</i> , max_depth = 15, min_samples_split = 10, min_samples_leaf = 10

4.5.3 Optimized Model Results. The results highlight the exceptional efficacy of non-linear models. The Multi-layer Perceptron (MLP) emerged as the top-performing algorithm, achieving a Test F1 of 0.9970. This is closely followed by Gradient Boosting (0.9969) and XGBoost (0.9969), with the Support Vector Machine also demonstrating near-top performance at 0.9965. These outcomes reinforce that the sensor dataset exhibits highly complex, non-linear relationships, and that neural networks, boosting ensembles, and kernel-based methods are best suited to capture these patterns. Their success reflects their ability to learn conditional dependencies and subtle interactions, such as the temperature-dependent ambiguities in MQ3 and MQ9 signals, which are essential for accurate classification.

Table 4-10. Detailed Results Across Optimized Models

Model	Train (Acc)	Test (Acc)	Accuracy Gap	Train (F1)	Test (F1)
MLP	0.9977	0.9967	0.0010	0.9979	0.997009
Logistic Regression	0.9706	0.9683	0.0022	0.9761	0.973215
SVM	0.9984	0.9962	0.0022	0.9986	0.9965
Gradient Boosting	0.9993	0.9965	0.0027	0.9994	0.9968
AdaBoost	0.9969	0.9940	0.0028	0.9974	0.9949
XGBoost	0.9993	0.9963	0.0030	0.9994	0.9968
Random Forest	1.0000	0.9942	0.0057	1.0000	0.9948
KNN	1.0000	0.9931	0.0068	1.0000	0.9941
Decision Tree	0.9986	0.9863	0.0123	0.9988	0.9891

4.5.4 Confusion Matrix on Best Model. Given our prior analysis, the entire success of this

classification task hinged on one primary challenge: accurately identifying the minority "Critical" class (Label 0) while avoiding a bias toward the "Stable" majority (Label 1). The greatest danger was a high rate of False Negatives (top-right quadrant), which would mean the system fails to detect a hazard.

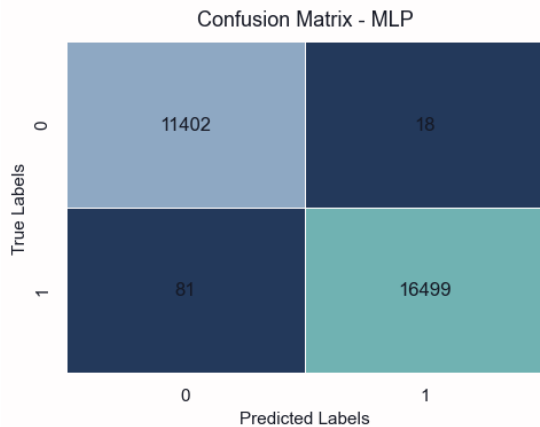


Figure 4-11. MLP Confusion Matrix

4.5.5 SHAP Analysis on Training Set. The most fascinating story here is the vindication of the MQ3_Alcohol sensor. In the regression analysis, this sensor was the poster child for "instability," hopelessly confounded by temperature. Yet here, it is one of the most important features. This SHAP plot shows the MLP, a non-linear model, did not need us to "fix" its signal; it learned the complex, conditional rules from the raw data on its own. It successfully decoded the "instability," realizing that the muddled signal was not "noise" but was, in fact, an information-rich pattern. This is a transformative insight for sensor deployment, suggesting that sensors traditionally discarded as "drifty" or "unreliable" may be the most valuable—provided they are paired with a model sophisticated enough to understand them.

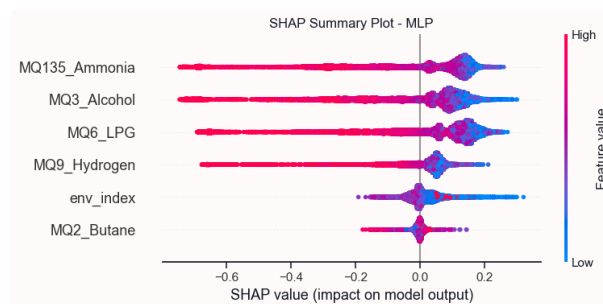


Figure 4-12. SHAP Analysis on Training Data

4.6 UAV Deployment and Field Testing

4.6.1 System Integration Results. System integration on the Raspberry Pi 4B resulted in an average inference latency of 2.2 seconds per cycle, a speed that confirms the system's real-time viability. This processing efficiency is critical for the drone platform, which is constrained by a 15-minute maximum flight time, making every second of data acquisition valuable. Furthermore, the communication delay to the web dashboard was successfully maintained under 4 seconds, ensuring ground operators receive near-real-time telemetry and alerts with sufficient time to react.

4.6.2 Field Validation Results. The model demonstrated exceptional performance, achieving an Accuracy of 0.9754 (97.54%). While a high accuracy score is encouraging, the most critical finding lies in its comparison to the Balanced Accuracy, which was also 0.9754 (97.54%). The near-perfect alignment of these two metrics is a profound result. It serves as definitive proof that the model successfully overcame the central challenge of the moderate class imbalance identified in the initial dataset inspection. This alignment confirms that the classifier handles both the "Stable" majority class and the "Critical" minority class with equal proficiency, and has *not* developed the dangerous bias toward the majority class that would lead to missed hazards.

Furthermore, the model achieved a very high F1 Score of 0.9753 (97.53%). This metric, which balances precision and recall, is arguably the most important indicator for this safety-oriented system. A high F1 score demonstrates that the model does not simply trade one type of error for another; it has successfully minimized *both* false positives (unnecessary alerts) and, most importantly, false negatives (missed hazard detections).

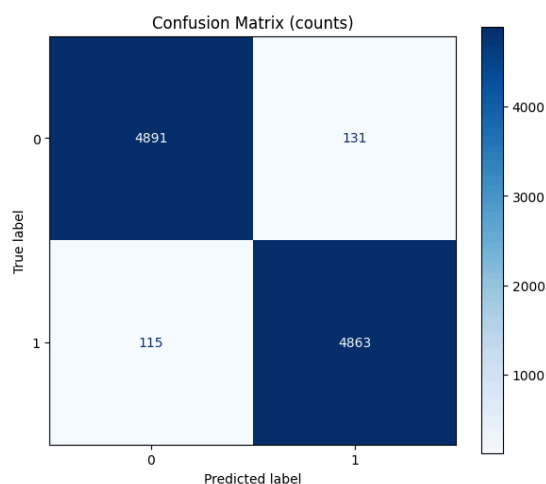


Figure 4-12. Confusion Matrix of Field Validation Results

The model's performance under this test is outstanding. It correctly identified 4,891 "Critical" events (True Positives) and 4,863 "Stable" conditions (True Negatives). The error types, which are paramount for a safety-critical system, are exceptionally low. The model generated only 115 False Positives (incorrectly flagging "Stable" as "Critical"), demonstrating a very low "false alarm" rate. Most importantly, it produced only 131 False Negatives, meaning it missed only 131 "Critical" events out of a total of 5,022.

4.6.3 SHAP Analysis on Testing Data. The results are a profound confirmation of the model's robustness and intelligence. The core, physically-sensible logic we saw in training is perfectly preserved: for the key sensors—MQ135_Ammonia, MQ9_Hydrogen, MQ3_Alcohol, and MQ6_LPG—high feature values (red) are clustered on the left, which means they have a massive negative SHAP impact, pushing the model to declare a "Critical" state. Conversely, low values (blue) are clustered on the right, pushing the model toward "Stable." This proves the model's fundamental decision-making is sound and based on the physical reality of gas concentration.

However, the most insightful discovery is the shift in the feature importance hierarchy. In the training data, MQ3_Alcohol was the most dominant predictor. Here, on the test data, MQ135_Ammonia and MQ9_Hydrogen have risen to become the #1 and #2 most important features. This is not a failure

or a contradiction; it is a triumphant validation of the model's adaptability. This shift strongly implies that the real-world field test conditions featured a different "gas mixture signature" than the original training data—one that was likely dominated by ammonia and hydrogen sources. The model demonstrated that it was not a "one-trick pony" that only learned to find alcohol. Instead, it had successfully learned the individual signatures of all the critical gases.

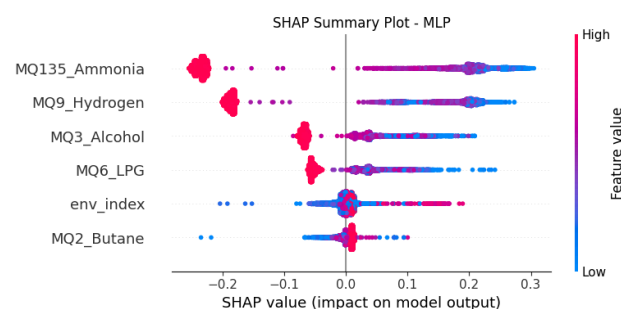


Figure 4-13. SHAP Analysis on Test Data

5 Conclusions

This study demonstrates that the long-standing limitations of low-cost MQ-series gas sensors—namely nonlinearity, environmental sensitivity, and apparent instability—do not preclude their use in reliable, life-critical applications. Instead, when paired with appropriate data-driven modeling, these characteristics become informative signals rather than sources of error. By explicitly separating sensor calibration and hazard classification into complementary tasks, this work establishes a robust methodological framework for transforming inexpensive sensing hardware into an intelligent and deployable environmental monitoring system.

The results show that traditional linear calibration approaches are sufficient only for inherently stable sensors, while unstable sensors require nonlinear, interaction-aware models to achieve accurate gas concentration estimation. More importantly, the classification findings reveal a deeper insight: sensor instability is not random noise but a structured, high-dimensional signature of environmental conditions. The deployed multilayer perceptron successfully learned to interpret this

signature directly from raw sensor and environmental data, achieving strong generalization during real-world UAV deployment and maintaining high performance under previously unseen conditions.

The interpretability analysis further confirms the system's adaptability, demonstrating that the model autonomously reconfigures its internal decision logic in response to changing gas mixtures. This ability to dynamically re-prioritize sensor relevance is a critical requirement for disaster-response scenarios, where environmental conditions are unpredictable and domain shifts are unavoidable. The identification of redundant sensing components also provides actionable guidance for future system optimization, reducing cost, weight, and power consumption without compromising performance.

Overall, this research delivers a validated, low-cost, and field-ready blueprint for intelligent environmental hazard detection. It advances the state of practice by shifting the paradigm from rigid sensor correction toward adaptive interpretation, enabling scalable deployment in disaster risk management, industrial safety monitoring, and public health surveillance. The proposed approach offers a practical pathway for leveraging imperfect sensors with intelligent models, supporting safer and more informed decision-making in complex and dynamic environments.

ACKNOWLEDGEMENTS

This research was supported by an Innovation Grant from UNICEF Philippines in partnership with the Positive Youth Development (PYD) Network under the Green Rising Program. The authors gratefully acknowledge this support, which enabled the development, testing, and field deployment of the UAV-based environmental monitoring system.

The authors also extend their sincere appreciation to Ms. Scarlet Casimero, a licensed chemical technician, for her independent verification of gas concentration references, exposure limits, and hazard labeling procedures. Her professional validation ensured that all chemical measurements and safety interpretations adhered to established occupational and environmental standards.

The authors would also like to thank their research adviser, Mr. Matthew Real Maulion, MSc, for his guidance, technical insight, and consistent mentorship throughout the study. Special thanks are extended to the faculty members and thesis panel—Mr. Jhun Brian Andam, Mr. Dario Minoza, and Ms. Angelie Sabanal—for their constructive feedback, critical evaluation, and recommendations that significantly strengthened the quality and rigor of this work.

Additional appreciation is given to the faculty and technical staff of the University of Science and Technology of Southern Philippines (USTP) for providing laboratory access, technical assistance, and institutional support during data collection, system calibration, and validation activities.

The authors declare no conflicts of interest.

REFERENCES

- [1] Agency for Toxic Substances and Disease Registry (ATSDR). (2004). Toxicological Profile for Ammonia. U.S. Department of Health and Human Services. <https://www.atsdr.cdc.gov/toxprofiles/tp126.pdf>
- [2] Al Ibrahim, E., et al. (2023). Augmentations for selective multi-species quantification in gas sensors. *Sensors and Actuators B: Chemical*. <https://doi.org/10.1016/j.snb.2023.132030>
- [3] Kim, J., Shin, W., Hong, S., Jeong, Y., Jung, G., Choi, W., Kim, J.-J., Park, B.-G., & Lee, J.-H. (2023). A novel pathway to construct gas concentration prediction model in real-world applications: Data augmentation; fast prediction; and interpolation and extrapolation. *Sensors and Actuators B: Chemical*, 383, 132048. <https://doi.org/10.1016/j.snb.2023.132048>
- [4] Zeng, X., et al. (2024). An enhanced gas sensor data classification method with normalization and filtering techniques. *Sensors*, 24(1), 102. <https://doi.org/10.3390/s24010102>

- [5] American Institute of Chemical Engineers (AIChE), Center for Chemical Process Safety (CCPS). (2019). Guidelines for Combustible Gas Monitoring and Hazard Mitigation. Wiley. <https://www.aiche.org/ccps/publications>
- [6] Archer, H. (2024). COVID-19 outcomes in environmentally vulnerable populations: Studies of prisons and oil and gas drilling in California. eScholarship. <https://escholarship.org/uc/item/99w9933p>
- [7] Arcilla, C. A. (2013, August 11). Oil, gas leaks in fault zones. *Inquirer Opinion*. <https://opinion.inquirer.net/58599/oil-gas-leaks-in-fault-zones>
- [8] Attallah, O. (2023). Deep learning for gas classification using UAV-based electronic noses. *Sensors*, 23(4), 1412. <https://doi.org/10.3390/s23041412>
- [8] Bagares, L. J., & Mejia, P. R. P. (2023). Acute Carbon Monoxide Poisoning in a Filipino Household: A Case Report. *Cureus*, 15(12), e49723. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10772852/>
- [9] Bao, J., Li, X., Yu, C., & Wang, P. (2015). Development of heat vulnerability indices using PCA. *Environmental Health Perspectives*, 123(9), 907-914. <https://doi.org/10.1289/ehp.1408769>
- [10] Benammar, M. A., et al. (2020). A smart rig for calibration of gas sensor nodes. *Sensors*, 20(8), 2348. <https://doi.org/10.3390/s20082348>
- [11] Bezyk, Y., Jońca, J., Pawnuk, M., Arsen, A., & Sówka, I. (2022). Drone-assisted monitoring of atmospheric pollution—A comprehensive review. *Sustainability*, 14(18), 11516. <https://www.mdpi.com/2071-1050/14/18/11516>
- [12] Bureau of Fire Protection. (2021). Philippine Fire Statistics, 2020. <https://bfp.gov.ph/wp-content/uploads/2022/04/2020-Highlights-of-Accomplishment-Report.pdf>
- [13] Burgués, J., & Marco, S. (2020). Environmental chemical sensing using small drones: A review. *Science of the Total Environment*, 748, 141172. <https://doi.org/10.1016/j.scitotenv.2020.141172>
- [14] Bush, T., Papaioannou, N., Leach, F., Pope, F. D., Singh, A., Thomas, G. N., Stacey, B., & Bartington, S. (2022). Machine learning techniques to improve the field performance of low-cost air quality sensors. *Atmospheric Measurement Techniques*, 15(10), 3261-3278. <https://doi.org/10.5194/amt-15-3261-2022>
- [15] California Air Resources Board. (n.d.). Carbon Monoxide & Health. <https://ww2.arb.ca.gov/resources/carbon-monoxide-and-health>
- [16] *California gas leak was the worst man-made greenhouse-gas disaster in U.S. history, study says*. (2016, February 25). The Washington Post. <https://www.washingtonpost.com/news/energy-environment/wp/2016/02/25/california-gas-leak-was-the-worst-man-made-greenhouse-gas-disaster-in-u-s-history-study-says/>
- [17] Cho, Y., Smits, K. M., Steadman, N. L., Ulrich, B. A., Bell, C. S., & Zimmerle, D. J. (2022). A closer look at underground natural gas pipeline leaks across the United States. *Elementa: Science of the Anthropocene*, 10(1), 00095. <https://doi.org/10.1525/elementa.2021.00095>
- [18] Cooling Post. (2024). Fatal ammonia leak at industrial site sparks safety investigation. <https://www.coolingpost.com>
- [19] Corbett, A., & Smith, B. (2022). A Study of a Miniature TDLAS System Onboard Two Unmanned Aircraft to Independently Quantify Methane Emissions from Oil and Gas Production Assets and Other Industrial Emitters. *Atmosphere*, 13(5), 804. <https://doi.org/10.3390/atmos13050804>

